THE UNIVERSITY *of* EDINBURGH

Development of a Modular *in Vivo* Reporter System for CRISPR-mediated Genome Editing and Its Therapeutic Applications for Rare Genetic Respiratory Diseases

THE UNIVERSITY *of* EDINBURGH

Robert Graham Foster

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2018

Preface

Signed Declaration

I hereby declare that:

(a) This thesis has been composed entirely by myself.
(b) This work is my own, and that if any other individual or member of my group has made a significant contribution to any of the work presented that this has been clearly indicated.
(c) This work has not been submitted for any other degree or professional qualification.

Robert Graham Foster

2nd August 2018

Acknowledgements

Undertaking a PhD is a mammoth endeavour, and one which I never would have been able to complete without the aid and support of very many people through the years. Here I would like to try and thank everyone who has helped me progress all the way from being a lowly undergraduate, to a PhD student on the verge of submission.

Firstly, I would like to thank my honours-project supervisors Dirk-Jan Kleinjan and Shipra Bhatia, in the exceptional lab of Veronica van Heyningen, for giving me my first real taste of research and providing such a supportive and friendly atmosphere when I was just starting in the lab. It was this experience which gave me the desire to do research, and lead to where I am today.

Secondly, I must thank Greg Kudla for taking me on post-undergraduate, to do a (paid!) bioinformatics project for which I also had very little experience, and teaching me all the ins and outs of bioinformatics work. These great experiences at the IGMM lead to me applying for, and being awarded, a PhD position here, where I have (quite amazingly) been for the past four-and-a-half years.

During the course of my PhD project I have been surrounded by so many supportive individuals it would be difficult to mention them all here, but I will try! To start with, I must thank Chris Boyd for initially taking me on after my rotations and giving me a huge amount of guidance which I still follow to this day. Furthermore, I must thank all the members of his lab for creating such a friendly and convivial atmosphere, along with all the members of the CGEM PhD office for the great conversation and outings during my time there.

Secondly, I must thank Pleasantine Mill for offering to be my co-supervisor following my first-year review, and providing me with a new, constructive direction for my project. Without her, my PhD would likely never have been as successful as it has been, and she has provided me with an enormous amount of support during the past few years, for which I am hugely grateful. Then I must also thank all the other members of her and Ian Jackson's lab, both past and present, who have provided me with so much help and training. In particular I must mention Girish Mali and Peter Tennant for all their help training me in various skills, along with all the wonderful support

staff such as Peter Budd and Margaret Keighren, without whom I would have no idea how to do cloning or mouse work. Furthermore, a big shout-out must go to all the members of my current office such as Amy, Dan and Pete for making it enjoyable to come in every day. And finally, I must note all the support I have received from technical services at the IGMM; especially Elisabeth Freyer for the huge amount of FACS work and simply being a great person to chat to, along with the imaging and bioinformatics teams for helping me with all the bizarre analysis questions I have had. The IGMM has truly been a wonderful place to undertake my PhD.

Abstract

Rare diseases, when considered as a whole, affect up to 7% of the population, which would represent 3.5 million individuals in the United Kingdom alone. However, while "personalised medicine" is now yielding remarkable results using recent sequencing technologies in terms of diagnosing genetic conditions, we have made much less headway in translating this patient information into therapies and effective treatments. Even with recent calls for greater research into personalised treatments for those affected by a rare disease, progress in this area is still severely lacking, in part due to the astronomical cost and time involved in bringing treatments to the clinic.

Gene correction using the recently-described genome editing technology CRISPR/Cas9, which allows precise editing of DNA, offers an exciting new avenue of treatment, if not cure, for rare diseases; up to 80% of which have a genetic component. This system allows the researcher to target any locus in the genome for cleavage with a short guide-RNA, as long as it precedes a highly ubiquitous NGG sequence motif. If a repair sequence is then also provided, such as a wild-type copy of the mutated gene, it can be incorporated by homology-directed repair (HDR), leading to gene correction. As both guide-RNA and repair template are easily generated, whilst the machinery for editing and delivery remain the same, this system could usher in the era of 'personalised medicine' and offer hope to those with rare genetic diseases. However, currently it is difficult to test the efficacy of CRISPR/Cas9 for gene correction, especially in vivo.

Therefore, in my PhD I have developed a novel fluorescent reporter system which provides a rapid, visual read-out of both non-homologous end joining (NHEJ) and homology-directed repair (HDR) driven by CRISPR/Cas9. This system is built upon a cassette which is stably and heterozygously integrated into a ubiquitously expressed locus in the mouse genome. This cassette contains a strong hybrid promoter driving expression of membrane-tagged tdTomato, followed by a strong stop sequence, and then membrane-tagged EGFP. Unedited, this system drives strong expression of membrane-tdTomato in all cell types in the embryo and adult mouse. However, following the addition of CRISPR/Cas9 components, and upon cleavage, the tdTomato is rapidly excised, resulting via NHEJ either in cells without fluorescence (due to

imperfect deletions) or with membrane-EGFP. If a repair template containing nuclear tagged-EGFP is also supplied, the editing machinery may then use the precise HDR pathway, which results in a rapid transition from membrane-tdTomato to nuclear-EGFP. Thereby this system allows the kinetics of editing to be visualised in real time and allows simple scoring of the proportion of cells which have been edited by NHEJ or corrected by HDR. It therefore provides a simple, fast and scalable manner to optimise reagents and protocols for gene correction by CRISPR/Cas9, especially compared to sequencing approaches, and will prove broadly useful to many researchers in the field.

Further to this, I have shown that methods which lead to gene correction in our reporter system are also able to partially repair mutations found in the disease-causing gene, Zmynd10; which is implicated in the respiratory disorder primary ciliary dyskinesia (PCD), for which there is no effective treatment. PCD is an autosomal-recessive rare disorder affecting motile cilia (MIM:244400), which results in impaired mucociliary clearance leading to neonatal respiratory distress and recurrent airway infections, often progressing to lung failure. Clinically, PCD is a chronic airway disease, similar to CF, with progressive deterioration of lung function and lower airway bacterial colonization. However, unlike CF which is monogenic, over 40 genes are known to cause PCD. The high genetic heterogeneity of this rare disease makes it well suited to such a genome editing strategy, which can be tailored for the correction of any mutated locus.

Lay Summary

While “personalised medicine” has recently come into its own and allowed clinicians to provide many patients suffering from rare genetic conditions with a diagnosis, we have made much less headway in translating this information into therapies and successful therapeutics. In the case of the rare respiratory disease primary ciliary dyskinesia (PCD), genetic mutations have been identified in nearly 40 genes and yet there is still no effective treatment for this debilitating condition; mostly due to the astronomical cost in bringing therapies to clinic. PCD patients suffer from a lifelong, severe and progressive lung disorder, which urgently requires the development of an effective and aggressive intervention. The recent game-changer for rare human genetic disease is the possibility of gene correction using targeted “molecular scissors” which allow precise and efficient editing of our DNA sequence. This powerful technology can accurately cut out the disease-causing sequence of DNA and paste a correct copy in its place. If this occurs in the right cell types, this technology offers the possibility of a cure for genetic diseases like PCD. The biggest obstacles to making this a reality are (a) targeted delivery of this machinery and (b) making it more efficient at repair.

To fast-track solutions to these problems, I have developed a fluorescent reporter mouse model that allows us to sensitively visualise gene-correction events within our bodies. Using this model we can quickly and accurately see under the microscope which cells we have corrected and how many. In this way, it will allow researchers to rapidly optimise approaches for gene correction in a live model, which was previously impossible. Furthermore, I have optimised a toolkit containing several types of gene editing reagents which can be delivered to any cell type in the body; which can themselves be grown in culture outside the reporter mouse. Using this reporter, I have determined that (1) the motile-ciliated cell types which are affected in PCD can be edited; (2) that the editing we see by fluorescence is an accurate read-out of changes in the DNA sequence and (3) that the reporter provides a rapid read-out of these genetic changes. Furthermore, I have also tested whether these gene-editing tools can be improved by adding small molecules which alter DNA repair pathways.

As a result of my thesis work, the lab is ready to test these optimised reagents in live reporter mice and their embryos, with the ultimate goal of using the tools to correct

the defects found in mice with PCD. It is our hope that our corrective gene-editing strategy for PCD would intervene early to limit the damage caused by recurrent infections, reduce medication use and decrease hospitalisations for children with PCD.

Contents

List of Figures

List of Tables

List of Abbreviations

%	Percent
*	Statistically significant to $p<0.05$
**	Statistically significant to $p<0.01$
***	Statistically significant to $p<0.001$
°C	Degrees Celsius
aa	Amino acid
AAV	Adeno-associated virus
ABT	ABT-751
ADA-SCID	Adenosine deaminase deficiency, severe combined immunodeficiency
AIFA	Agenzia Italiana del Farmaco (Italian Medicines Agency)
amp	Ampicillin
AmpR	Ampicllin resistance gene
ATP	Adenosine triphosphate
B6	C57BL/6J Mouse Strain (Black6)
BALB/cJ	Inbred, albino laboratory mouse strain
BFA	Brefeldin A, putative HDR enhancer
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BRCT	BRCA1 C terminus
CAR	Chimeric antigen receptor
Cas9	CRISPR-associated protein 9
Cas9n	Cas9 nickase
cat. no.	Catalogue number
CDK	Cyclin-dependent kinase
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CMOS	Complementary metal-oxide semiconductor
CMV	*Cytomegalovirus*
CRISPR	Clustered regularly interspersed short palindromic repeats

crRNA	CRISPR RNA
DAPI	4',6-diamidino-2-phenylindole
dCas9	Inactive spCas9
ddH_2O	Double-distilled water
del	Deletion
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
DNAAF	Dynein Axonemal Assembly Factor
DNAH	Dynein axonemal heavy chain
DNAI	Dynein axonemal intermediate chain
DNALI	Dynein axonemal light intermediate chain
DNA-PK	DNA-dependent protein kinase
DNA-PKcs	DNA–protein kinase catalytic subunit
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
DSB	Double strand break
DSBR	Double-strand break repair
DsRed	Red flourescent protein derived from *Discosoma*
E	Embryonic day
E. coli	*Escherichia coli*
EDTA	Ethylene diamine tetra acetic acid
EGFP	Enhanced green fluorescent protein
ES	Embryonic stem (cells)
EtBr	Ethidium bromide
EU	European Union
EURORDIS	European Organisation for Rare Diseases
F1	Cross between B6 and BALB/cJ
FACS	Fluorescence-activated cell sorting
FC	Flow cytometry
FCS	Fetal calf serum
FDA	US Food and Drug Administration

Fiji	Distribution of ImageJ with included analysis packages
FOXJ1	Forkhead box J 1
g	Gram
gDNA	Genomic DNA
GFP	Green fluorescent protein
GOI	Gene of interest
gRNA	Guide RNA
H2B	Histone H2B
HBoV1	Human bocavirus type 1
HBSS	Hank's Balanced Salt Solution
HBV	Hepatitis B virus
HDR	Homology directed repair
HEK293	Human embryonic kidney cell line
HEK293T	HEK293 line stably expressing SV40 large T antigen
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HITI	Homology-independent targeted integration
HIV	Human immunodeficiency virus
hPSC	Human pluripotent stem cell
HR	Homologous recombination
Hz	Hertz
IGMM	Institute of Genetics and Molecular Medicine
Indel	Insertion or deletion
iPSC	Induced pluripotent stem cell
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISFET	Ion-sensitive field-effect transistor
IVT	*In vitro* transcription
kb	Kilobase
KI	Knock-in
l	Litre
L-755,507	β3-adrenergic receptor partial agonist
LB	Luria-Bertani (broth)
LRS	Long Read Sequencing

Lv	Lentivirus
M	Molar
m	*Prefix* membrane-tagged
m	*Prefix* milli
min	Minute(s)
MARCKS	Myristoylated alanine-rich C-kinase substrate
MEF	Mouse embryonic fibroblast
MMEJ	Microhomology-mediated end joining
MRN	MRE11-RAD50-NBS1 complex
MTEC	Mouse tracheal epithelial cell
mTmG	Membrane-Tomato, membrane-EGFP
n	*Prefix* nano
NF-H_2O	Nuclease-free water
NGS	Next generation sequencing
NHBE	Normal human primary bronchial epithelial cells
NHEJ	Non-homologous end joining
NHS	National Health Service
NLS	Nuclear localisation signal
ns	Not significant
nt	Nucleotide
NU-7441	Selective DNA-PK inhibitor, putative HDR enhancer
ODA	Outer dynein arm
OH	Hydroxyl group
ORPK	Oak Ridge Polycystic Kidney
p	Probability
P/S	Penicillin/streptomycin
pA/polyA	Polyadenylation signal
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
pCA	Synthetic enhancer-promoter-splice acceptor construct
PCD	Primary ciliary dyskinesia
PCR	Polymerase chain reaction

PI3K	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PNK	Polynucleotide kinase
pre-crRNA	Pre-CRISPR RNA
PTC	Premature termination codon
QLB	Quick Lysis Buffer
RNA	Ribonucleic acid
RNAi	RNA interference
RNP	Ribonucleoprotein
RP	Retinitis pigmentosa
RPA	Replication protein A
RS-1	RAD51-stimulatory compound 1
S. pyogenes	*Streptococcus pyogenes*
s/sec	Second(s)
SaCas9	Cas9 orthologue from *Staphylococcus aureus*
SAP	Shrimp alkaline phosphatase
SCR7	DNA ligase IV inhibitor
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDSA	Synthesis-dependent strand annealing
sgRNA	Single guide RNA
SIN	Self-inactivating
siRNA	Short interfering RNA
SMRT	Single-molecule real-time sequencing
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SPAdes	St. Petersburg genome assembler
spCas9	Cas9 from *S. pyogenes*
SRS	Short Read Sequencing
SS	Serum-starved/Serum-starvation
SSA	Single-strand annealing
SSB	Single strand break
ssDNA	Single-stranded DNA

SSN	Site-specific nuclease
ssODN	Single-stranded oligodeoxynucleotide
SV40	Simian vacuolating virus 40
T4 PNK	T4 polynucleotide kinase
TALEN	Transcription activator-like effector nuclease
TBE	Tris/Borate/EDTA buffer
tdTomato	Tandem dimer tomato (DsRed fluorescent protein variant)
TE	Tris-EDTA (buffer)
Tm	Annealing temperature
TOPO	Toposiomerase-based
tracrRNA	Trans-activating crRNA
Tris	Tris(hydroxymethyl)aminomethane
US	United States of America
UTR	Untranslated region
UV	Ultraviolet
V	Volt
VSV-G	Vesicular stomatitis virus glycoprotein G
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WT	Wild type
WTCRF	Wellcome Trust Clinical Research Facility
x g	Times gravity
X-Gal	5-Bromo-4-Chloro-3 Indolyl-Beta-D-thiogalactopyranoside
ZFN	Zinc-finger nuclease
ZMW	Zero mode waveguide
Zmynd10/ZMYND10	Zinc finger MYND-type containing 10 (gene)
Zmynd10/ZMYND10	Zinc finger MYND-type containing 10 (protein)
ZnF	Zinc finger
µ	*Prefix* micro

Chapter 1 Introduction - The Therapeutic Application of Genome Editing for Rare Genetic Respiratory Diseases

1.1 Rare Diseases

Rare diseases are defined by their limited life-time prevalence in a population, with various international bodies outlining their own parameters for classification. In the United States, for example, a disease is considered rare when it affects less than 200,000 individuals (U.S. Food and Drug Administration 2017), whereas in the European Union (EU), a rare disease is seen to affect less than 1 in 2,000 citizens (European Organisation for Rare Diseases 2005). Despite this variation, as over 5,800 rare diseases are currently listed on the web portal Orphanet (www.orpha.net/, accessed 05/02/2018), considered as a whole, they will come to affect a large proportion of the population. Indeed, in the EU in 2005, it was estimated approximately 30 million individuals suffered from a rare disease, or 6 to 8% of the total population (European Organisation for Rare Diseases 2005).

Patients suffering from a rare disease are disadvantaged compared to sufferers of more common conditions in several ways. Firstly, there is often a significant delay between a patient initially presenting to a medical practitioner and them receiving a correct diagnosis, often on the scale of years. This is partly due to a lack of knowledge and awareness, but also since the presence of many of these conditions can only be confirmed after genetic testing (Pogue, Cavalcanti et al. 2018). Indeed, up to 80% of rare diseases are believed to have a genetic component (European Organisation for Rare Diseases 2005), and 3573 genes are listed as associated with rare diseases on Orphanet (www.orpha.net/, accessed 08/02/2018). Recent advances in sequencing technologies have partially improved diagnoses for patients but have also lead to a substantial increase in the number of genes implicated each year. Even with a correct diagnosis, given the rarity of each condition, there may be very few physicians available with the skills to treat the condition of interest (Pogue, Cavalcanti et al. 2018). Treatment options are then further limited as pharmaceutical companies are generally unwilling to invest in rare disease drug development, due to the high costs and small market size involved. Currently, 400 drugs are licensed in the US for treating

rare conditions, covering only 5% of the rare disease burden (Zhao and Wei 2017), and each of these treatments is often prohibitively expensive (European Organisation for Rare Diseases 2005).

Given the severely limited therapeutic options for rare diseases, governmental bodies have rallied behind improved access to current treatments and the development of novel therapeutics (Gammie, Lu et al. 2015). In particular, the United Kingdom (UK) has published a comprehensive strategy for rare diseases. This strategy sets ten goals which are to be achieved by 2020, including the establishment of specialised clinical centres for care, the promotion of research into rare diseases, including collaboration with the NHS (National Health Service), and the rapid translation of advances in understanding through novel infrastructure. Furthermore, this strategy recognises the impact next-generation sequencing (NGS) technologies will have on reducing the cost of genetic diagnosis, and in identifying novel causative genes. As such, it recommends the adoption of NGS into mainstream diagnosis pathways by 2020 (UK Department of Health 2013).

1.2 Next-Generation Sequencing and Genetic Diagnosis

Prior to the development of NGS technologies, genetic diagnosis was performed on a gene-by-gene basis using Sanger sequencing, affording a very limited power of detection, especially for conditions associated with multiple genes. However, with advancements in sequencing capacities, it became possible to simultaneously sequence several disease-causing genes in a 'gene panel', and then later perform whole exome sequencing (WES) relatively inexpensively (Caspar, Dubacher et al. 2017), with the potential to detect 85% of known disease-causing mutations (Choi, Scholl et al. 2009). However, the current gold-standard for diagnosis is whole genome sequencing (WGS), which has been shown to more accurately detect copy number variations (CNVs), indels and single nucleotide variants (SNVs) (Soden, Saunders et al. 2014, Meienberg, Bruggmann et al. 2016, Stavropoulos, Merico et al. 2016), as well as affording the possibility of detecting mutations in non-exonic, regulatory regions. While the average price of WGS is still relatively high, it is rapidly falling and expected to be below $5,000 per patient by 2020, although advancements in data analysis will need to be

made to prevent this becoming a time and cost limitation (Weymann, Laskin et al. 2017).

Currently, the majority of WES and WGS tests are carried out using short-read sequencing (SRS) techniques. In SRS, the input DNA is randomly sheared into 100-400bp fragments followed by adapter ligation and massively parallel sequencing. The current market leader in this field is Illumina, with Ion Torrent (Thermo Fisher) being a significant competitor (Caspar, Dubacher et al. 2017), both using variations of a 'sequencing by synthesis' approach (Goodwin, McPherson et al. 2016). In Illumina sequencing, reversible terminators are employed to allow the sequential addition of single, fluorescently-tagged nucleotides which can be detected by laser illumination. Following detection, the fluorescent dyes are removed, creating a new 3'-OH and allowing the addition of the next tagged nucleotide. In repeating these steps, a whole fragment can be sequenced with high accuracy (Figure 1.1A) (Ambardar, Gupta et al. 2016). In Ion Torrent sequencing, adapter-ligated fragments are fixed in micro-wells and then nucleotide species are sequentially washed over the wells. Upon incorporation of a base, an H^+ ion is released resulting in a change in pH which is detected by a complementary metal-oxide semiconductor (CMOS) and an ion-sensitive field-effect transistor (ISFET) device (Figure 1.1B) (Goodwin, McPherson et al. 2016). One disadvantage of this approach is its imprecision in quantifying the number of bases in homopolymer stretches, based on the intensity of the pH signal, leading to a characteristic error profile (Ambardar, Gupta et al. 2016).

More recently, third-generation, or long-read sequencing (LRS), approaches have been described which may overcome the issues SRS has in detecting variants in repetitive regions, due to the challenge of short read alignment (Caspar, Dubacher et al. 2017). There are currently two main players in the LRS field, PacBio with their single-molecule real-time sequencing (SMRT) approach, and Oxford Nanopore Technologies with single-molecule nanopore sequencing. In SMRT, a high-fidelity DNA polymerase is immobilised in a zero-mode waveguide (ZMW), a microscopic pore which allows the visualisation of individual fluorescent nucleotides. The DNA polymerase is then specifically illuminated, allowing detection of fluorescently-labelled nucleotides as they are incorporated (Figure 1.2A) (Ambardar, Gupta et al. 2016). This method can produce reads of up to 10kb, although with an ~10% error rate

(Nagarajan and Pop 2013). The alternative platform, the Oxford Nanopore MinION, makes use of a proprietary DNA helicase which 'feeds' single-stranded DNA through a proprietary protein pore which is fixed in a polymer membrane. A current is passed through the membrane, and as each base travels through the pore its unique shape creates a characteristic disturbance in the current which can be recorded (Ambardar, Gupta et al. 2016) (Figure 1.2B). Currently, MinION reads have a similar length distribution and slightly higher error rate than SMRT reads (Lu, Giordano et al. 2016), although MinION sequencing is significantly easier to perform, with minimal sample preparation and the only equipment required being a USB-sized device (Ambardar, Gupta et al. 2016). Planned future updates to the MinION system should significantly reduce the error rate and increase throughput (Next Gen Seek 2016), which may potentially lead to a paradigm shift in NGS, allowing WGS in the space of a few hours on the bench-top (Jain, Olsen et al. 2016).

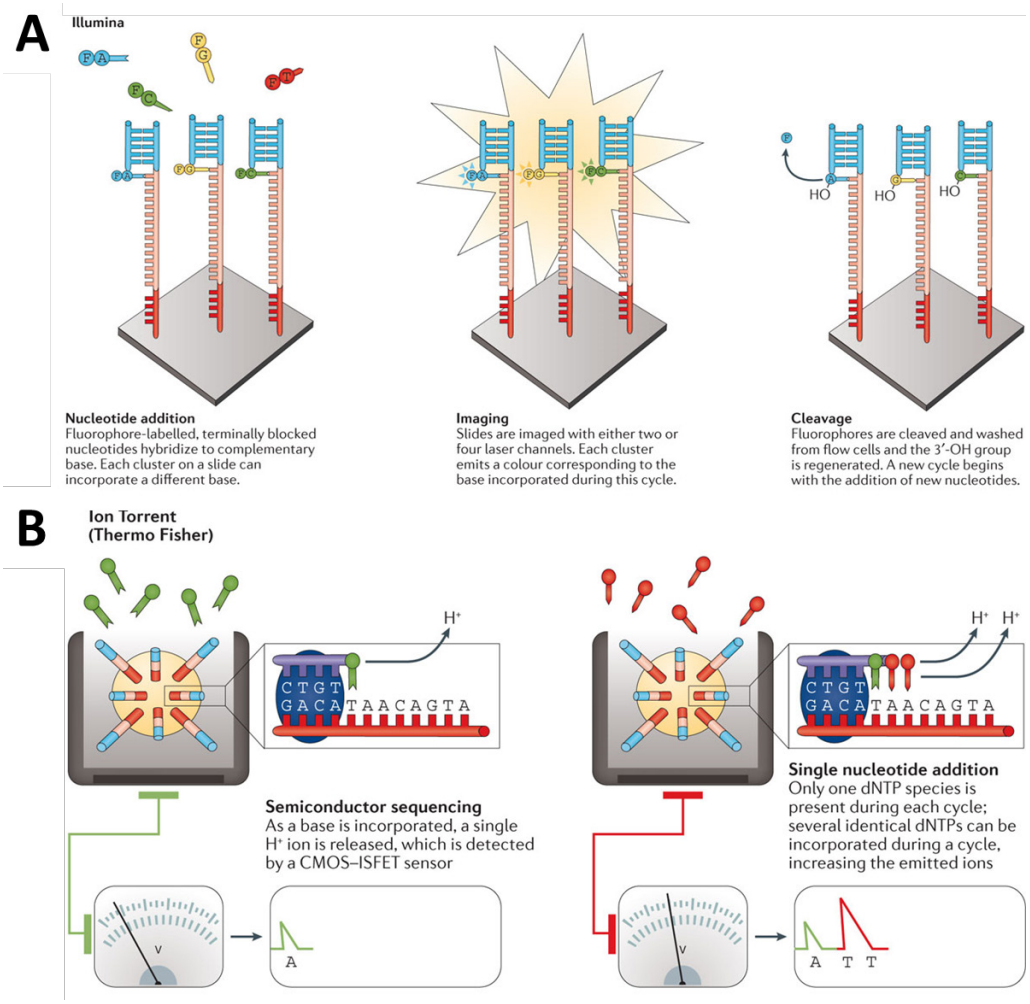


Figure 1.1 (on previous page): Comparison of short-read sequencing technologies from Illumina and Thermo Fisher (Ion Torrent). A) Illumina: Templates are hybridised to slides and then amplified. Primers and DNA polymerase are added to the flow cell. Terminally blocked, fluorescently-labelled nucleotides are then added, resulting in the incorporation of a single base. Unincorporated bases are removed by washing and the slide is illuminated. The fluorescent dye is then cleaved, regenerating the 3'-OH and allowing the addition of another terminally blocked nucleotide in the next cycle. B) Ion Torrent: Templates are PCR amplified on beads, which are placed individually into micro-wells. Nucleotide species are sequentially added to the wells. Upon base incorporation, the H^+ release results in a change in pH which is detected by an integrated complementary metal-oxide semiconductor (CMOS) and an ion-sensitive field-effect transistor (ISFET) device. Adapted from Goodwin, McPherson et al. (2016).

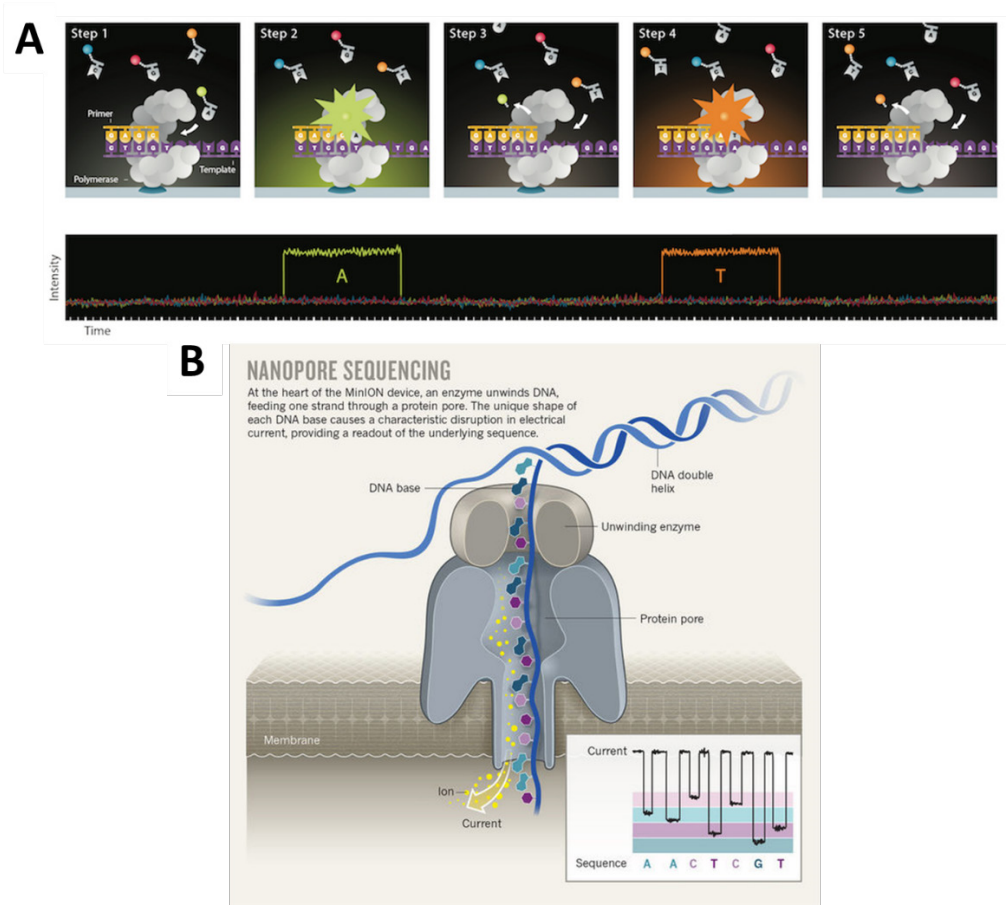


Figure 1.2: Comparison of long-read sequencing (LRS) technologies. A) SMRT sequencing: 1) fluorescently-tagged nucleotides are added to the ZMW; 2) the DNA polymerase immobilises the incorporated base in the detection volume; 3) the attached fluorescent dye is released upon phosphate chain cleavage; and 4–5) the process is repeated. Taken from Clinical Lab Products (2017). B) Nanopore sequencing: The unwinding enzyme passes the single-stranded DNA through a protein pore in a charged membrane. Disturbances in the membrane caused by the translocation of nucleotides result in a characteristic change in current which is detected (insert). Taken from Eisenstein (2017).

1.3 Gene Therapy

With the advancements in NGS technologies, we will continue to detect ever greater numbers of disease-causing mutations. Gene therapy, using various vectors to transfer a wild-type copy of a mutated gene into affected cells, offers the potential to permanently functionally correct mutations that do not have a dominant-negative effect, as opposed to drug treatments which require repeated dosing (Dunbar, High et al. 2018). Furthermore, a gene therapy approach will be applicable to all mutations found in a gene, whereas the efficacy of drugs will likely depend on the exact mutation.

As nucleic acids are anionic polymers, they cannot interact with negatively charged plasma membranes, and as such require synthetic vectors to allow delivery into cells (Mottais, Le Gall et al. 2017). A substantial amount of research has therefore been dedicated to identifying and optimising delivery vectors for gene therapy, with the ideal vector often depending on the gene and tissue of interest. As the subject of this thesis is rare respiratory diseases, I shall focus on gene therapy of the lung.

Much of the work done on gene delivery to the pulmonary epithelium has been focussed on the rare, autosomal-recessive disease cystic fibrosis (CF), which is caused by mutations in a single gene, the cystic fibrosis transmembrane conductance regulator (*CFTR*). Defects in the chloride channel CFTR lead to the production of thick, sticky mucus in the airways and obstructive lung disease which is fatal in ~80% of cases, with an average life expectancy of 37 years (Cutting 2014). Recent clinical trials have demonstrated success using the liposome GL67A and plasmid pGM169 in delivering *CFTR* to the airways, although the improvement in lung function was very modest (Alton, Armstrong et al. 2015). The airways have a complex, branching structure, and in order to reach the cell layer, therapeutics also have to penetrate a layer of mucus and the periciliary layer (PCL) (Figure 1.3). Furthermore, in a disease state such as cystic fibrosis, the mucus layer is substantially thickened, and may contain interacting particles such as cellular debris, DNA and bacteria, which will all reduce efficiencies (Mottais, Le Gall et al. 2017). Therefore, it may be necessary to move to a viral vector approach to achieve high transgene expression, exploiting the natural ability of viruses to enter their host cells (Castellani and Conese 2010).

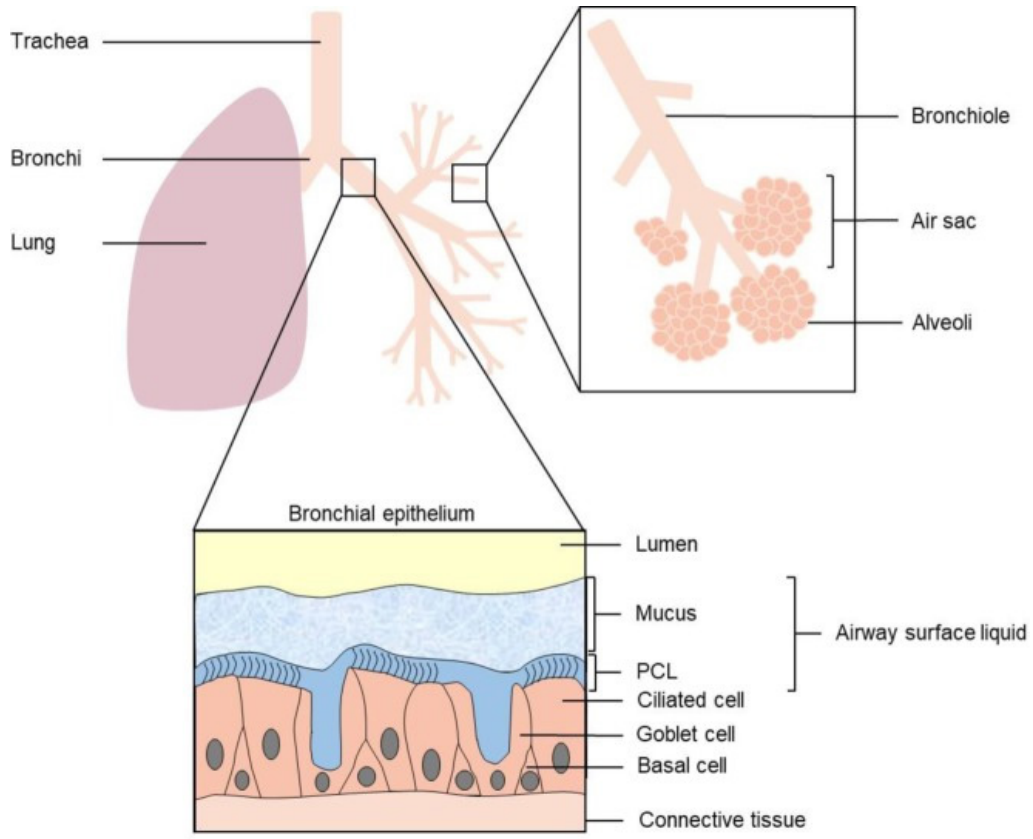


Figure 1.3: Diagrammatic representation of the human airway, showing the branching structure present. Within the airway, target cells are covered by the airway surface liquid. Taken from Mottais, Le Gall et al. (2017).

The two virus types which currently show the most promise in gene therapy of the lung are lentivirus, specifically the HIV-1 strain, and adeno-associated virus (AAV). Lentiviral vectors have a packaging capacity of ~8kb, are able to transduce non-dividing cells, and provide efficient transgene expression (Castellani and Conese 2010). Furthermore, adaptations to produce self-inactivating (SIN), or replication-incompetent lentiviral vectors, have made them largely bio-safe (Miyoshi, Blömer et al. 1998, Zufferey, Dull et al. 1998). Unfortunately, whilst their safety profile is good, most pseudotypes display a limited ability to transduce a fully differentiated epithelium, unless agents that disrupt tight-junctions are simultaneously applied (Castellani and Conese 2010), although they may be effective in a neonatal or fetal context (Buckley, Howe et al. 2008). To overcome this barrier, it has been shown that efficient transduction of the respiratory epithelium via the apical side without conditioning can be achieved by pseudotyping the Simian Immunodeficiency Virus (SIV) with envelope proteins (F and HN) from Sendai virus to produce F/HN-SIV

(Mitomo, Griesenbach et al. 2010, Griesenbach, Inoue et al. 2012). Importantly, F/HN-SIV was shown to be permissive to repeat administration in the mouse, with increased expression depending on dose (approximately twice that of the non-viral formulation) and no discernible toxicity. Furthermore, it gave persistent expression in human lung slices and airway epithelial cells (Griesenbach, Inoue et al. 2012). This vector has thence been shown to drive expression in ~15% of the target epithelial cells in mouse (c.f. with the 5-10% correction believed to be required to reverse the CF defect (Castellani and Conese 2010)), with no apparent effect from pre-existing immunity, and therefore has recently been recommended for progression into human trials (Alton, Beekman et al. 2017).

Alongside lentivirus, much current research is directed at the use of AAV vectors. AAV is naturally non-pathogenic in humans, replication-defective and largely non-integrating, which makes it attractive for gene therapy approaches, although one drawback is its small packaging capacity, at ~4.7kb, which limits its use for larger genes (Dunbar, High et al. 2018). However, as demonstrated with *CFTR* (Zhang, Wang et al. 1998, Wang, Fischer et al. 1999), it may be possible to produce truncated versions of genes which still drive high expression levels. AAV exists in nine naturally occurring serotypes, of which AAV6 has been shown to drive the highest levels of transgene expression in the murine respiratory epithelium and normal human primary bronchial epithelial cells (NHBE) (Kurosaki, Uchibori et al. 2017). Moreover, AAV genomes and capsid proteins can be recombined; the combination AAV2/5 (AAV2 genome and AAV5 capsid proteins) has been discovered to also drive high levels of expression in human intestinal organoids and the mouse lung (Vidović, Carlon et al. 2015). Further increases may be seen in efficiencies by directed evolution of the capsid proteins, as has shown promise in the CF pig (*Sus domesticus*) (Steines, Dickey et al. 2016). Alternatively, the AAV genome can also be packaged in other capsids, such as that of human bocavirus type 1 (HBoV1). The HBoV1 capsid offers a larger packaging capacity of ~5.5kb, while also efficiently transducing the human airway epithelium via the apical membrane (Yan, Feng et al. 2017). However, work still needs to be done in translating these recent advances clinically for the treatment of diseases such as CF (Guggino and Cebotaru 2017).

1.4 Gene Correction by Site-Specific Nucleases

As an alternative to the above gene therapy by gene transfer approaches, the discovery and characterisation of several site-specific nucleases (SSNs) has afforded the novel possibility of directly correcting mutated genes *in situ*, eliminating the chance of insertional mutagenesis and leaving a gene which will be controlled by its native promoters and enhancers. In addition, this approach may show particular promise in treating diseases caused by dominant negative mutations, such as *HTT* in Huntington's disease, which are un-targetable by gene addition, due to the efficiency at which these tools can 'knock-out' their target genes (Cox, Platt et al. 2015). Furthermore, it circumvents the limited packaging capacities of viral vectors, allowing the targeting of genes, and diseases, which were previously unfeasible.

SSNs are targeted to a specific sequence in the genome, whereupon their nuclease domains create a double-strand break (DSB). This DSB is generally repaired by one of two major cellular DSB repair pathways; non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ is template-independent and error-prone, allowing the creation of mutations and 'knock-outs', whereas HDR depends on the presence of a repair template and can elicit precise, bespoke changes (Cox, Platt et al. 2015) (Figure 1.4). Repair pathways are discussed further in Section 1.6.

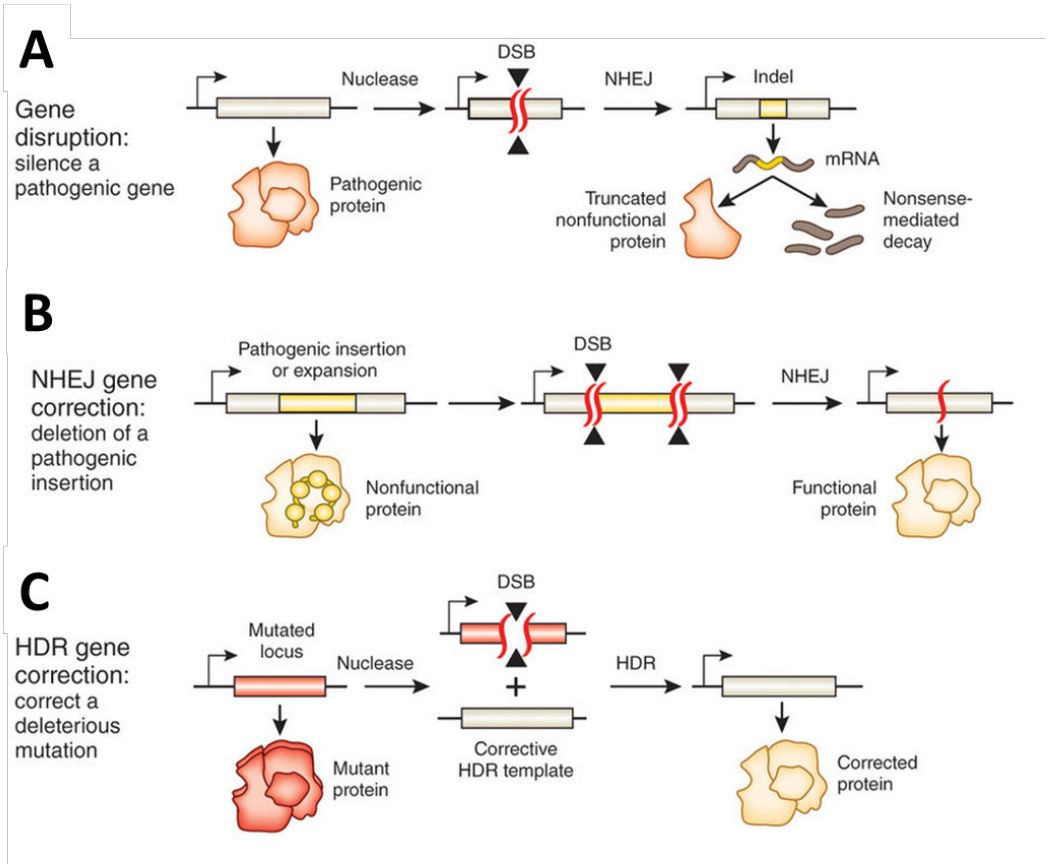


Figure 1.4: The application of site-specific nucleases for therapeutic genome editing. A) An SSN is targeted to a disease-causing gene/mutation, where it creates a DSB. This DSB is repaired by NHEJ, resulting in the formation of an indel and the production of a non-functional/degraded protein. B) An insertion is targeted on either side by two SSNs, resulting in cleavage which removes the fragment and the resulting locus is repaired by NHEJ. C) If a corrective repair template is supplied alongside an SSN, it may be incorporated by the HDR pathway resulting in the production of a correct gene copy ('gene correction'). Adapted from Cox, Platt et al. (2015).

The first SSN system described was based on the use of an array of zinc-fingers (ZnFs), a common DNA-triplet-binding motif found in eukaryotes, which was linked to the bacterial endonuclease FokI to create zinc-finger nucleases (ZFNs). While ZFNs were specific and efficacious, their uptake was limited due to the requirement to re-engineer and validate each array, a costly and time-consuming process. Following ZFNs, a new modality emerged upon the discovery of transcription activator-like effectors (TALEs). TALEs are formed of 33-35bp repetitive DNA-binding domains which each recognise a single nucleotide. Unlike ZFNs, the DNA-binding domains of TALEs are modular and can be easily combined to target any sequence in the genome. Furthermore, a nuclease can be attached to the TALE to form a TALEN, allowing

directed cleavage. As both approaches can be designed to require a significant amount of homologous sequence (>30bp) for cleavage, they are highly specific in the human genome (Gaj, Gersbach et al. 2013) (Figure 1.5). Furthermore, both ZFNs and TALENs have shown promise in clinical translation. ZNFs have recently been used to render CD4 T cells resistant to HIV infection by mutating its receptor *CCR5*. On infusion of CD4 T cells back into patients, levels of HIV dropped and in one case became undetectable (Tebas, Stein et al. 2014). TALENs have been employed in a similar *ex vivo* approach, and used to autologously edit T cells to express a chimeric antigen receptor (CAR) against the leukaemia antigen CD19. Upon re-infusion into two patients, marked remissions were observed (Qasim, Zhan et al. 2017). However, the therapeutic application of SSN-mediated genome editing is still in its infancy.

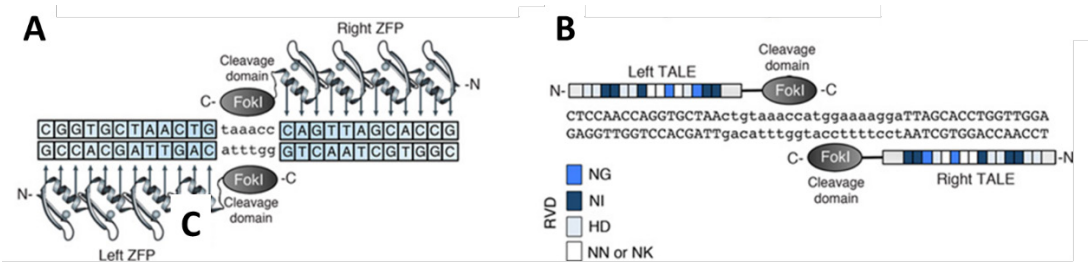


Figure 1.5: Diagrammatic representation of zinc-finger nuclease (ZFN) and TALEN dimers bound to DNA. A) ZFNs: left and right zinc-finger proteins (ZFPs) are designed to target a 5-7bp sequence recognised by the FokI cleavage domain. B) 33-35bp TALE repeats bind in a base-specific manner due to two repeat-variable di-residues (RVDs). Left and right TALEN target sites are separated by a spacer sequence (12-20bp). RVD compositions are indicated. Adapted from Gaj, Gersbach et al. (2013).

1.5 The CRISPR/Cas9 System

While ZFNs and TALENs have shown promise in therapeutic genome editing, their use has been rapidly superseded by the recently-described CRISPR/Cas9 system for genome editing. CRISPR, short for clustered regularly interspaced short palindromic repeats, was originally used to describe functionally obscure repetitive arrays which were found in the genomes of most prokaryotes. These arrays were invariably found to associate with four CRISPR-associated, or Cas proteins, which were absent in CRISPR-negative species (Jansen, Embden et al. 2002). Later, it was found that these arrays contained invasive DNA sequences termed ‘spacers’ (Bolotin, Quinquis et al. 2005, Mojica, Diez-Villasenor et al. 2005), that these spacers could be gained upon

phage adaptation (Pourcel, Salvignol et al. 2005, Barrangou, Fremaux et al. 2007), and that they were required, together with the Cas proteins, for phage resistance (Barrangou, Fremaux et al. 2007). Therein, the CRISPR/Cas system was found to be a form of prokaryotic immune system. Soon after, it was found that DNA is the target of CRISPR/Cas (Marraffini and Sontheimer 2008), that target sites are always followed by a conserved protospacer adjacent motif (PAM) (Mojica, Diez-Villasenor et al. 2009) and that Cas9 is an endonuclease which cleaves DNA precisely 3nt from the 5' end of the PAM (Garneau, Dupuis et al. 2010).

Today, many different CRISPR/Cas systems have been described, although the best characterised is the CRISPR/Cas9 system from *Streptococcus pyogenes* (spCas9). It was shown that the non-coding CRISPR RNAs (crRNAs) required by spCas9 for cleavage could be combined into a (single) guide-RNA (sgRNA/gRNA), and that these gRNAs could be employed to direct Cas9 cleavage *in vitro* (Jinek, Chylinski et al. 2012). Soon after this discovery, two groups demonstrated that this system could be adapted for efficient, RNA-guided genome editing in mammalian cells (Cong, Ran et al. 2013, Mali, Yang et al. 2013), and that all the components necessary for editing could be delivered effectively on a single plasmid, with the only requirement being to clone in a specific 20bp 'target sequence' (Mali, Yang et al. 2013, Ran, Hsu et al. 2013). The ease-of-use and low cost of this system, combined with its comparable or higher efficiency than TALENs (Cong, Ran et al. 2013, Mali, Yang et al. 2013), has led to an explosion of publications in the genome editing field in recent years (Stella and Montoya 2016).

1.5.1 Off-Target Editing with the CRISPR/Cas9 System

The CRISPR/Cas9 system functions similarly to ZFNs and TALENs, in that it creates a DSB which is repaired by the cellular machinery, resulting in NHEJ or HDR (Figure 1.4). However, unlike the other two systems, Cas9 creates a DSB on binding only one strand with a 20nt target sequence (Figure 1.6A), of which it initially appeared only 8-12bp was required for cleavage (Jinek, Chylinski et al. 2012, Cong, Ran et al. 2013). This would afford much less specificity than is seen with ZFNs or TALENs, raising fears of off-target cleavage and mutagenesis. Subsequently, a high rate of off-target mutagenesis was indeed detected in human cells with spCas9 (Fu, Foden et al. 2013,

Hsu, Scott et al. 2013). To address this issue, a 'nickase' version of Cas9 (Cas9n) was developed (Cong, Ran et al. 2013) which needs to be targeted to two adjacent target sites to produce a DSB (Figure 1.6B). This approach was found to offer similar on-target editing efficiencies to wild-type Cas9, in terms of NHEJ and HDR, with 50- to 1000-fold reduced off-target mutagenesis (Ran, Hsu et al. 2013). However, this approach also severely limits the number of potential on-target sites in the genome.

Recently, a significant amount of research has been carried out to address the issue of off-target mutagenesis. Several online tools have been developed to bioinformatically predict off-target sites in the genome, such as GuideScan (Perez, Pritykin et al. 2017), CRISPOR (Haeussler, Schönig et al. 2016) and Cas-OFFinder (Bae, Park et al. 2014), allowing guides with likely off-targets to be avoided. Furthermore, a wide array of sequencing approaches have been performed to attempt to quantitate and characterise the off-target profile of Cas9, including targeted sequencing based on off-target predictions (Fu, Foden et al. 2013, Hsu, Scott et al. 2013), ChIP-seq to sequence regions bound by Cas9 (Duan, Lu et al. 2014, Kuscu, Arslan et al. 2014, Wu, Scott et al. 2014, O'Geen, Henry et al. 2015), whole genome sequencing (Smith, Gore et al. 2014, Veres, Gosis et al. 2014), and 'unbiased' approaches for mapping and sequencing DSBs (Crosetto, Mitra et al. 2013, Tsai, Zheng et al. 2014, Kim, Bae et al. 2015). Whilst varied, generally these efforts have shown that the rates of off-target mutagenesis driven by Cas9 are substantially lower than feared with most guides (Duan, Lu et al. 2014, Smith, Gore et al. 2014, Veres, Gosis et al. 2014, Wu, Scott et al. 2014, Kim, Bae et al. 2015, O'Geen, Henry et al. 2015), and that recent algorithms allow accurate prediction of off-target sites with mutation rates over 0.1% (Bae, Park et al. 2014, Haeussler, Schönig et al. 2016). Moreover, a highly sensitive sequencing tool failed to detect mutations in the vast majority of off-target sites with 4 or greater mismatches (Tsai, Zheng et al. 2014).

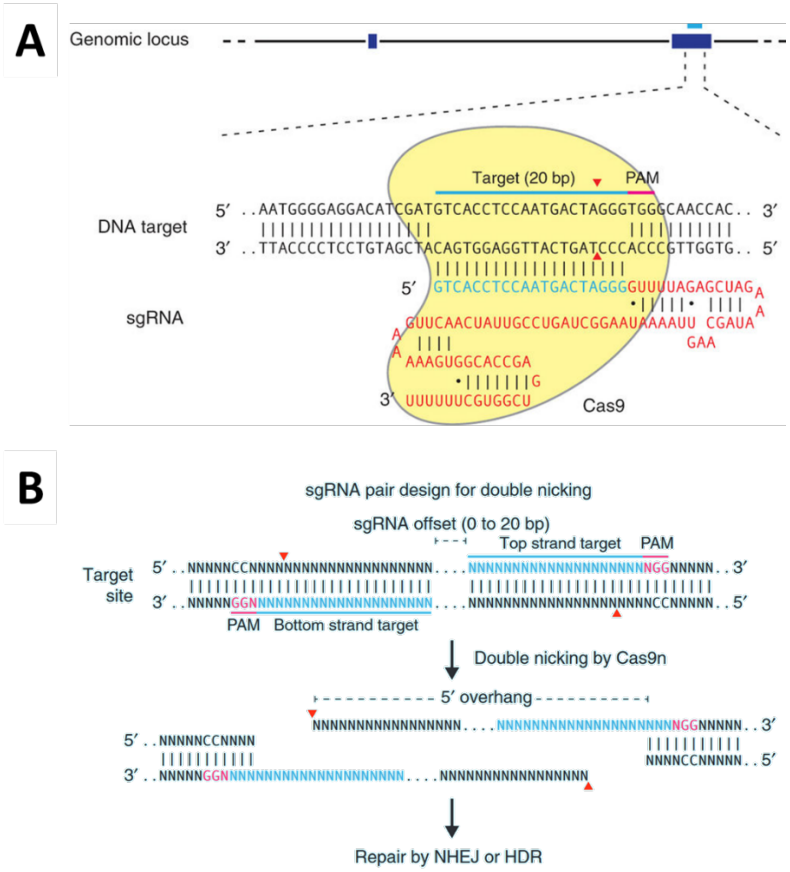


Figure 1.6: Schematic representation of the wild-type Cas9 (A) and Cas9-nickase 'double nicking' (B) strategies for editing DNA. A) Cas9: a single 20bp guide-RNA (sgRNA) directs double strand cleavage by the Cas9 endonuclease, 3 bases 5' of the PAM (pink arrowheads). B) Cas9n double nicking: 20nt target sites are chosen on both strands, separated by a 0-20bp offset. Nicking at both sites by Cas9n leads to the creation of a DSB with 5' overhangs. Adapted from Ran, Hsu et al. (2013).

Whist the rates of off-target mutagenesis are therefore very low, and essentially negligible in the context of a molecular biology lab (Veres, Gosis et al. 2014), even very infrequent events must be avoided in a clinical setting, where mutations risk leading to oncogenesis. Several strategies have been employed to reduce the frequency of off-target mutagenesis. Promising results have been reported with the use of truncated gRNAs, less than 20nt long, in providing a several-fold increase in targeting specificity (depending on the locus) with similar on-target editing (Fu, Sander et al. 2014, Tsai, Zheng et al. 2014, Zhang, Li et al. 2016). Furthermore, the Cas9 nuclease has been engineered to produce several higher specificity variants (Bolukbasi, Gupta et al. 2015, Slaymaker, Gao et al. 2015, Kleinstiver, Pattanayak et al. 2016, Chen, Dagdas et al. 2017), the most recent of which, evoCas9, being reported to increase

specificity 79-fold while maintaining 90% on-target efficiency (Casini, Olivieri et al. 2018). In addition, on-target specificity can be improved simply by titrating the quantity of Cas9 (Ran, Hsu et al. 2013), by limiting the duration of exposure through an inducible system (Cao, Wu et al. 2016) or by direct delivery of Cas9/gRNA complexes to cells which are degraded within 24 hours (DeWitt, Corn et al. 2017). Moreover, it is likely that these approaches could be combined, resulting in a system which is highly specific and with a very limited risk of off-target effects.

1.5.2 On-Target Editing with the CRISPR/Cas9 System

As with off-target editing, much work has been done in elucidating the factors which lead to high on-target editing efficiencies. A study by Doench, Hartenian et al. (2014), in which they screened a pool of gRNAs for their bi-allelic knock-out ability, revealed that gRNAs which target the coding sequence, especially the exact exon-intron boundary, are highly favoured. Furthermore, they found specific nucleotide preferences within the protospacer, PAM and PAM-proximal sequences (Figure 1.7).

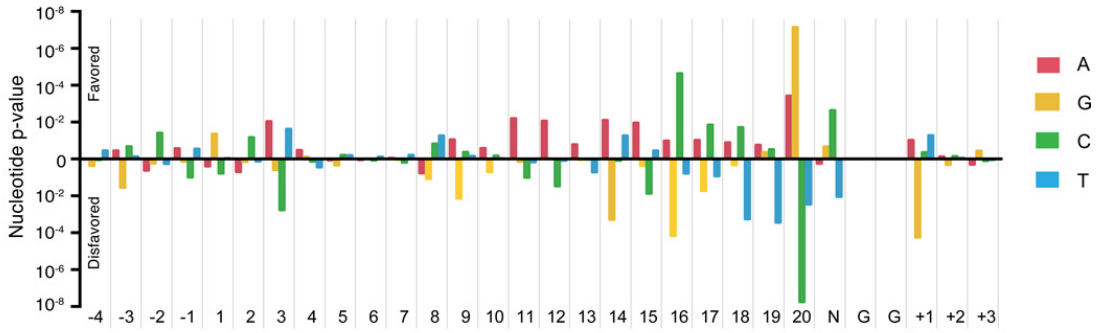


Figure 1.7: Nucleotide preference varies along the guide, PAM and adjacent sequences. Graph shows the p-values of observing a guide with a percent-rank activity of >0.8 at every position, based on 1,841 gRNAs targeting the CDS of 9 mouse genes. P-values were calculated from the binomial distribution from all coding-sequence targeting guides employed. Taken from Doench, Hartenian et al. (2014).

Besides sequence, it is hypothesised that chromatin context may play an important role in determining guide activities. Chromatin exists in two broad states, euchromatin and heterochromatin. Euchromatin is open, accessible for transcription and associated with active genes, whereas heterochromatin is densely packed and inaccessible (Watts 2016). Studies have found that editing by CRISPR/Cas9 is impeded in a heterochromatic context (Chen, Rinsma et al. 2016, Jensen, Floe et al. 2017), and that

NHEJ may be more affected than HDR (Chen, Liu et al. 2017). Furthermore, all mammalian DNA is packaged in nucleosomes, with some regions being more nucleosome-dense than others. A significant correlation has been observed between higher nucleosome-density and reduced binding of Cas9, although the effects of this may be locus specific due to the potential for nucleosome remodelling (Horlbeck, Witkowsky et al. 2016, Isaac, Jiang et al. 2016). Interestingly, an inactive, 'dead' version of Cas9 (dCas9) has been shown to induce an open chromatin state (Barkal, Srinivasan et al. 2016), a feature which may also find therapeutic application.

1.5.3 Delivery Vectors for CRISPR/Cas9-Mediated Genome Editing In Vivo

Whilst many simple and effective methods exist for employing the Cas9 system in cell culture systems (Luo and Saltzman 2000), especially plasmid (Ran, Hsu et al. 2013) and RNP transfection (DeWitt, Corn et al. 2017), achieving *in vivo* genome editing in a tissue of interest is a more complex task. A range of approaches can be deployed, the most simple of which being hydrodynamic injection of plasmid DNA, which has shown success in targeting several tissues in rodents, in particular hepatocytes (Suda and Liu 2007). Using CRISPR/Cas9 plasmids, this approach has been used to suppress hepatitis B virus (HBV) in mice (Lin, Yang et al. 2014, Ramanan, Shlomai et al. 2015, Zhen, Hua et al. 2015), although currently the method is not translatable to humans due to the risk of adverse cardiac events (Suda and Liu 2007).

An alternative method, also commonly employed in animal models, involves the direct injection of Cas9 mRNA and gRNA (with or without repair template) into oocytes or zygotes, and has shown high efficiencies in the mouse (Li, Qiu et al. 2013, Wang, Yang et al. 2013), rat (Li, Qiu et al. 2013) and zebrafish (Hruscha, Krawitz et al. 2013, Hwang, Fu et al. 2013), amongst other systems. Interestingly, a proof-of-principle report used this technique to correct a mouse model with a dominant mutation that causes cataracts (Wu, Liang et al. 2013). Alternatively, Cas9 ribonucleoprotein complexes (RNPs) can be delivered by injection (Sung, Kim et al. 2014), or electroporation (Chen, Lee et al. 2016), with similarly high efficiencies. However, germline targeting in humans remains the subject of intense ethical debate, although

controversial reports are already emerging of gene correction in human zygotes (Ma, Marti-Gutierrez et al. 2017, Tang, Zeng et al. 2017).

For gene correction approaches to prove successful in humans outside of the embryo, it may be necessary to move to a viral delivery approach, as with conventional gene therapy (Lau and Suh 2017). AAV vectors are also considered to be the most promising for CRISPR/Cas9 gene editing, although due to the ~4.2kb size of spCas9, it has been necessary to employ minimal promoters and poly-adenylation signals to stay within packaging limits (Senís, Fatouros et al. 2014). Alternatively, a recent study took advantage of the split-intein protein trans-splicing technique (Li, Sun et al. 2008), to deliver the spCas9 gene on two separate AAV vectors, freeing >2kb on each vector for additional elements and retaining high efficiencies (Chew, Tabebordbar et al. 2016). Recently, several smaller spCas9 alternatives have been described (Mir, Edraki et al. 2018), the most notable of which being *Staphylococcus aureus* Cas9 (saCas9), which is smaller in size (~3.2kb) but has a more restrictive PAM sequence (Friedland, Baral et al. 2015). SpCas9 and saCas9 AAV-based methods have already demonstrated success in several tissues of the mouse, including muscle (Nelson, Hakim et al. 2016, Amoasii, Long et al. 2017, Bengtsson, Hall et al. 2017), heart (Xie, Zhang et al. 2016, Amoasii, Long et al. 2017, El Refaey, Xu et al. 2017, Ishizu, Higo et al. 2017) and retina (Yu, Mookherjee et al. 2017).

In addition, it is necessary to consider the format of delivery of the repair template. Whilst single-stranded oligonucleotides (ssODNs) may be effective in eliciting small changes in tissues amenable to HDR such as the embryo (Hwang, Fu et al. 2013, Wang, Yang et al. 2013, Wu, Liang et al. 2013), larger templates such as plasmids may be required for longer insertions (Yang, Wang et al. 2013). In tissues less amenable to HDR, more efficient repair templates may be required. Minicircle plasmid derivatives may offer greater repair efficiencies, potentially due to their lack of bacterial sequences which can elicit an immune response (Kay, He et al. 2010). Furthermore, the inability of the immune system to readily detect minicircles may increase their residency times and allow them to be repeatedly dosed (Munye, Tagalakis et al. 2016), leading to higher total rates of HDR. Alternatively, AAV itself appears to have an innate ability to stimulate HDR, along with a high packaging capacity (Gaj, Epstein et al. 2015), although repeated dosing may not be possible (Beck, Jones et al. 1999, Chirmule, Xiao

et al. 2000). Interestingly, a ten-fold increase in HDR has been noted when using an RNP-Cas9 and AAV repair approach over a dual plasmid strategy in a HEK293T reporter cell line (Gaj, Staahl et al. 2017).

1.6 Double-Strand Break Repair Mechanisms

Once a double-strand break has been created by an SSN, the resulting editing outcome is largely decided by the cellular machinery. In mammalian cells there are two major DSB repair pathways, NHEJ and homologous recombination (HR, employed for HDR), and several minor repair pathways, the most important for SSN-mediated mutagenesis being microhomology-mediated end-joining (MMEJ) and single-strand annealing (SSA) (Figure 1.8A). The choice of repair pathway depends to a large extent on the cell's position in the cell cycle (Figure 1.8B). As there is no endogenous repair template available in mammalian cells outside of S/G2 phase (when replicated DNA is present), mammalian cells largely rely on the template independent, error-prone NHEJ pathway to repair DSBs, which are common and can arise due to various factors such as oxidative damage and ionising radiation (Lieber 2010).

In NHEJ, upon DSB creation, the proteins Ku70/Ku80 bind the free ends forming Ku-DNA complexes (Figure 1.8A i) (Franco, Murphy et al. 2008). These Ku-DNA complexes are able to recruit nuclease, polymerase and ligase in any order, resulting in a large range of potential indels (Lieber 2008). Further cleavage is mediated by the DNA–protein kinase catalytic subunit (DNA-PKcs) which binds to Ku-DNA and activates the endo/exonuclease Artemis (Franco, Murphy et al. 2008). In addition, the template-independent DNA polymerase mu, as well as DNA polymerase lambda, are able to bind Ku-DNA via their BRCT (BRCA1 C terminus) domains. Lastly, the complex XLF-XRCC4-DNA ligase IV is required to ligate the free ends (Ahnesorg, Smith et al. 2006), whether or not cleavage or extension has occurred. DNA ligase IV is non-specific and able to ligate incompatible ends across substantial gaps.

During the S and G2 phases of the cell cycle, the alternative repair pathways, HR, MMEJ and SSA become available. All three pathways depend on an initial 5' to 3' resection event mediated by CtIP/MRE11-RAD50-NBS1 (MRN) (Kakarougkas and Jeggo 2014, Seol, Shim et al. 2017), a complex which is regulated during the cell cycle by the presence of cyclin-dependent kinase (CDK) (Sfeir and Symington 2015).

Following resection, if there are small (~2-20bp) regions of homology present, the MMEJ pathway (Figure 1.8A ii) may be employed to repair the break, whereas larger regions of homology invoke the HR pathway, including the SSA pathway (Figure 1.8A iii) for resolving larger repeats (>14bp). Both MMEJ and SSA are highly mutagenic, resulting in deletions (MMEJ & SSA) and insertions (MMEJ) (Sfeir and Symington 2015). They consist of a similar process, including annealing, flap trimming, DNA synthesis and ligation, however MMEJ appears not to require the ssDNA-binding proteins RAD51/RAD52 or Ku70/Ku80 for repair (McVey and Lee 2008, Sfeir and Symington 2015). MMEJ instead appears to require replication protein A (RPA) for ssDNA-binding, and indeed the transition in occupancy of ssDNA ends between RPA and RAD52 at ~14bp appears to mediate the choice of repair pathway between MMEJ and SSA (Sfeir and Symington 2015).

In HR the ssDNA tail is also rapidly bound by RPA, however, unlike in MMEJ, RPA is then displaced by RAD51 in a BRCA2-dependent process. If no nearby intrachromatid homology is found, RAD51 then directs strand invasion of the homologous region of the sister chromatid and formation of a D-loop (Figure 1.8A iv) (Kakarougkas and Jeggo 2014). Following this initial formation of a D-loop, several different pathways can be employed to repair the break, including synthesis-dependent strand annealing (SDSA) and double-strand break repair (DSBR) (Li and Heyer 2008), the mechanisms governing which are poorly understood in mammalian (human) cells, but likely involve the interplay of several proteins and protein complexes (Kowalczykowski 2015). It is probable the relative frequency of MMEJ/SSA versus non-SSA HR is dependent on the presence of the sister chromatid for repair within a time-frame following DSB formation, as end-resection by CtIP/MRN can proceed automatically in S/G2 (Bhargava, Onyango et al. 2016).

Given this high complexity, the exact pathway employed following SSN-mediated cleavage will likely depend on several factors, including the nature of the locus and the nature of the repair template, if supplied. However, it must also be noted that Ku70 and Ku80 are still active during S/G2 and compete with CtIP for DNA binding, resulting in greater utilisation of the NHEJ pathway in mammalian cells even during this phase of the cell cycle (Kakarougkas and Jeggo 2014).

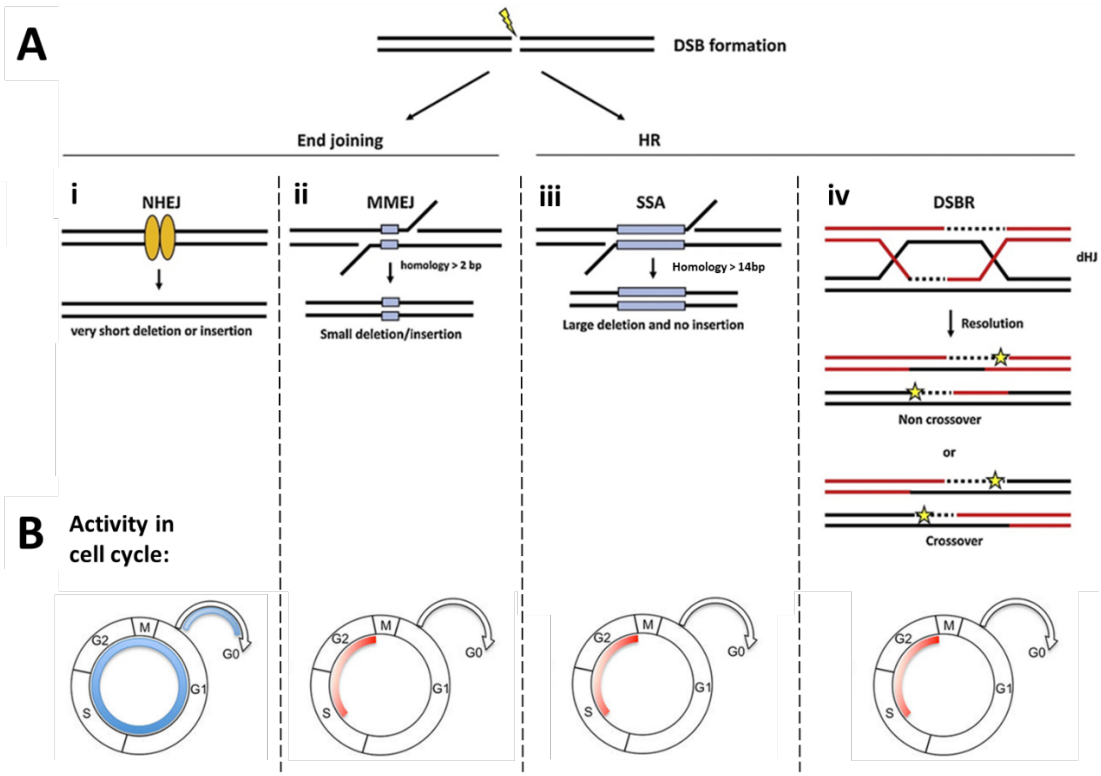


Figure 1.8: Schematic of the major repair pathways in mammalian cells, and their activity in the cell cycle. A) i) non-homologous end joining (NHEJ), yellow ovals indicate Ku-DNA and associated proteins; ii) micro-homology mediated end joining (MMEJ); iii) single-strand annealing (SSA); iv) double-strand break repair (DSBR), yellow stars indicate potential mutagenesis introduced by low-fidelity DNA synthesis. Light blue boxes indicate regions of homology. Sister chromatids are distinguished by colour. Dashed lines indicate newly-synthesised DNA. Adapted from Seol, Shim et al. (2017). B) Diagrammatic representation of the activity of each repair pathway in relation to the cell cycle. Adapted from Suzuki and Izpisua Belmonte (2018).

1.6.1 Small Molecule Enhancement of HDR

Given that the desired outcome of gene correction in mammalian cells will depend on the action of the infrequently-employed HR pathway, several studies have been performed to identify small molecules which can improve rates of HDR. As HR is active during S/G2 phases, trials have shown that small molecules which block the cell cycle in one of these phases can improve HDR. Nocodazole and ABT-751 (ABT) both block cells at the G2/M transition and have been shown to produce a 3-5 fold increase in HDR in hPSCs and iPSCs (Yang, Scavuzzo et al. 2016). Another study confirmed the increase in HDR on nocodazole treatment in HEK293T cells, human primary neonatal fibroblasts and human embryonic stem cells, and also found a similar if reduced effect with aphidicolin, which blocks in early S phase (Lin, Staahl et al. 2014).

Alternatively, small molecules have been identified which can perturb the NHEJ and HR pathways, leading to increases in HDR. SCR7 (pyrazine), an inhibitor of DNA ligase IV (Srivastava, Nambiar et al. 2012), has proven effective in a large array of cell lines (Chu, Weber et al. 2015, Maruyama, Dougan et al. 2015, Gkotzamanidou, Terpos et al. 2016, Li, Zhang et al. 2017, Shao, Ren et al. 2017, Hu, Shi et al. 2018), as has the DNA-PKcs inhibitor NU-7441 in HEK293 cells, MEFs and iPSCs (Robert, Barbeau et al. 2015, Zhang, Li et al. 2017). Alternatively, the HR-stimulator RS-1 (RAD51-stimulatory compound 1) also appears effective in HEK293A, U2OS and rabbit embryos (Pinder, Salsman et al. 2015, Song, Yang et al. 2016). Interestingly, two other compounds have also been discovered through screens to increase rates of HDR, through unexplained mechanisms, namely L-755,507 in many cell types (Yu, Liu et al. 2015, Li, Zhang et al. 2017) and BFA in mouse ES cells and iPSCs (Yu, Liu et al. 2015, Zhang, Li et al. 2017).

1.6.2 Alternatives to HDR

As these small molecule treatments may not be applicable in all cell types, especially *in vivo* and therapeutically, recently a new method has emerged which takes advantage of the NHEJ pathway to introduce a corrected fragment. Termed homology-independent targeted integration (HITI), it requires simultaneous Cas9-directed cleavage of an exogenous plasmid and the genomic target locus (Suzuki, Tsunekawa et al. 2016). The genomic target site is added in reverse direction to the plasmid/minicircle such that only correct insertion of the fragment will abolish further cleavage, leading to the production of virtually no clones with insertions in the reverse direction (Figure 1.9). Notably, this technique achieved several-fold higher rates of knock-in (KI) than an HDR donor, up to 60% in HEK293 cells. Critically, using AAV delivery in mice, they achieved KI rates of 3.5% in the brain following injection of the visual cortex, and KI rates of 4.2%, 3.4% and 10.0% in the liver, heart and quadriceps muscle respectively after systemic injection. Furthermore, they demonstrated this method can deliver substantial improvements in a rat model of retinitis pigmentosa (RP) following subretinal injections 3 weeks after birth (Suzuki, Tsunekawa et al. 2016). This targeted integration approach for gene therapy should partially address fears associated with insertional mutagenesis and afford somewhat better regulation of

the inserted gene, which could be introduced into its native locus, although, compared with HDR, care must be taken not to disrupt or introduce regulatory elements.

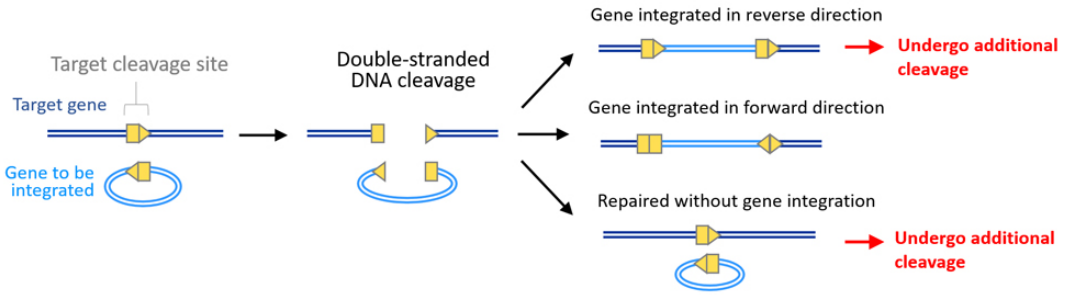


Figure 1.9: Schematic representation of the homology-independent targeted integration (HITI) system. Blue rectangles depict gRNA binding sites and blue triangles their associated PAM sites. Correct insertion of the gene of interest (GOI) results in the disruption of the target site, whereas reverse integration does not, leading to re-cleavage. Taken from Riken Center for Developmental Biology (2017).

Alternatively, novel Cas9 fusion proteins have recently been produced which allow 'base editing' within a defined window, typically 5nt, following Cas9 binding to DNA. Taking advantage of the rat cytidine deaminase APOBEC1 to elicit a C to T change (Komor, Kim et al. 2016), or a highly engineered RNA adenosine deaminase to elicit an A to G change (Gaudelli, Komor et al. 2017), these 'base editors' allow efficient (~35-50%) and targeted conversion of point mutations which are causative of human disease, with low rates of indel mutations (~0.1-1.1%, Komor, Kim et al. (2016), Gaudelli, Komor et al. (2017). They therefore provide an attractive alternative to HDR and HITI where only single base-pair changes are required. One concern with this base editing strategy is the unwanted, and potentially deleterious, conversion of surrounding bases, however, recent work has shown that these proteins can be engineered to have narrower editing windows (Kim, Komor et al. 2017), alleviating these concerns if a nearby gRNA target sequence is present.

1.7 Project Rationale and Aims

As explained in the preceding sections, the ease and precision of genome editing afforded by the CRISPR/Cas9 system opens an exciting new therapeutic avenue for the treatment, and potentially cure, of many genetic diseases. By creating bespoke guides and repair templates, which can be delivered alongside universal editing machinery using standardised protocols and delivery vectors, this 'modular' approach

allows truly personalised genetic medicine. With improved genetic diagnosis facilitated by the advancement of next-generation sequencing (NGS) technologies, we have the opportunity to develop tailored gene therapies for many rare genetic diseases, which would not have previously been deemed economically viable targets. However, for this to be a feasible therapeutic option, two significant hurdles must be overcome:

a) improving delivery of the CRISPR components to the somatic tissues of interest *in vivo*; and
b) biasing editing events in the targeted cells towards repair and gene correction.

When I began my PhD, genome editing events were being monitored at a sequence-based level, creating a bottleneck in determining which editing events are occurring in which cell types. No suitable methods were available which would allow us to systematically address these two issues, therefore we identified the need to develop a system which:

1) allows rapid and robust scoring of CRISPR-mediated events, particularly rates of NHEJ and HDR;
2) is functional *in vivo;*
3) is functional in all cell types; and
4) is functional throughout the cell cycle.

In order to achieve this, we decided to repurpose a mouse fluorescent-based reporter model of Cre-mediated recombination, noting the similarities between this system and deletions instigated by CRISPR/Cas9, as well as the speed and ease of detecting changes in fluorescence as a read-out via microscopy and flow cytometry (FC) or FACS (fluorescence-activated cell sorting). This model will be discussed further in Chapters 3 and 4. Furthermore, we wished to validate the applicability of results generated with our reporter system to other, disease-causing loci. To do so, as a proof-of-principle, I began with a mouse model of the rare, genetic, respiratory disease primary ciliary dyskinesia (PCD), which is discussed in Chapter 5.

Chapter 2 Materials and Methods

2.1 Design of CRISPR Guide-RNAs

Guides were designed against the desired target sequences using an array of well-known online tools and were chosen based on their high scores, indicating high efficiency and specificity, across the three tools used. These tools are located at http://crispr.mit.edu/ (Zhang Lab 2017), http://zifit.partners.org/ZiFiT/ (Sander, Zaback et al. 2007, Sander, Maeder et al. 2010, Hwang, Fu et al. 2013, Zinc Finger Consortium 2017) and https://www.atum.bio/eCommerce/cas9/input (ATUM 2018), the latter formerly being known as DNA2.0. Guides were designed to cut as close to the desired targets as possible, that being the -7bp deletion in *Zmynd10* or the loxP sites in the *mTmG* cassette, although guides which exactly overlapped these regions were not found. In the case of the *mTmG* system, the two guides used were both designed by another member of the lab, Peter Budd, whereas the three guides for *Zmynd10* were chosen by myself.

2.1.1 *mTmG* Locus

As stated above, the two guides designed against the loxP sequence in the *mTmG* system were both designed by Peter Budd using DNA2.0 (now ATUM), MIT CRISPR Design and further checked using ZiFit. The design of the guides takes advantage of the fact that there is 62bp of identical sequence surrounding both loxP sites, as based on the following BLASTn alignment:

Upper sequence: nt 2786 to 3019 **5' loxP site at 2886 to 2929**
Lower sequence: nt 5192 to 5425 **3' loxP site at 5292 to 5325**

```
EMBOSS_001         1 --CTGGGCAACGTGCTGGTTATTGTGCTGT-----CTCATC-----ATTT     38
                       |||     |.|.||.|||.|.|| .|||     ||||||     ||.|
EMBOSS_001         1 CACTG-----CATTCTAGTTGTGGT-TTGTCCAAACTCATCAATGTATCT     44

EMBOSS_001        39 TGGCAAAGAATTGATTTGATACCGCGGGCCCTCGACACT----AGTGAAC     84
                     |..||      ||..|.|||          || |||| |    ||| ||.
EMBOSS_001        45 TATCA------TGTCTGGAT----------CT-GACA-TGGTAAGT-AAG     75

EMBOSS_001        85 CT----CT-----TCGAGGGACCTAATAACTTCGTATAGCATACATTATA    125
                     ||    ||     ||||||||||||||||||||||||||||||||||
EMBOSS_001        76 CTTGGGCTGCAGGTCGAGGGACCTAATAACTTCGTATAGCATACATTATA    125

EMBOSS_001       126 CGAAGTTATATTAAGGGTTCCGGATC---------CATGGGTTGCTGTTT    166
                     |||||||||||||||||||||||||         .|||||||||||||
EMBOSS_001       126 CGAAGTTATATTAAGGGTTCCGGATCATCACCGCGGATGGGTTGCTGTTT    175

EMBOSS_001       167 CTCCAAGACCATGGTG-AGC-AAGGGAGAGGAGGTCATCAAAGAGTTC--    212
                     |||||||||     | ||| ||||||||.|..| |..|..||||..|
EMBOSS_001       176 CTCCAAGACC-----GCAGCGAAGGGAGAAGCCG-CCGCGGAGAGGCCTG    219

EMBOSS_001       213 ---ATGCGCTTCAAGGTGCGCATGG    234
                        |.|||      |.||    |||
EMBOSS_001       220 GGGAGGCG------GCTG----TGG    234
```

Figure 2.1: A BLASTn alignment of 100bp 5' and 3' of both loxP sites reveals 62bp of identical sequence. Here, the loxP sites are highlighted in green. Nucleotide positions are relative to the entire Rosa26 *mTmG* Sequence which can be accessed at https://www.addgene.org/17787/ (Muzumdar, Tasic et al. 2007).

In this 62bp of identical sequence, DNA2.0 identified three potential guide-RNAs, whereas MIT CRISPR Design identified four, however the top two guides identified by both are the same:

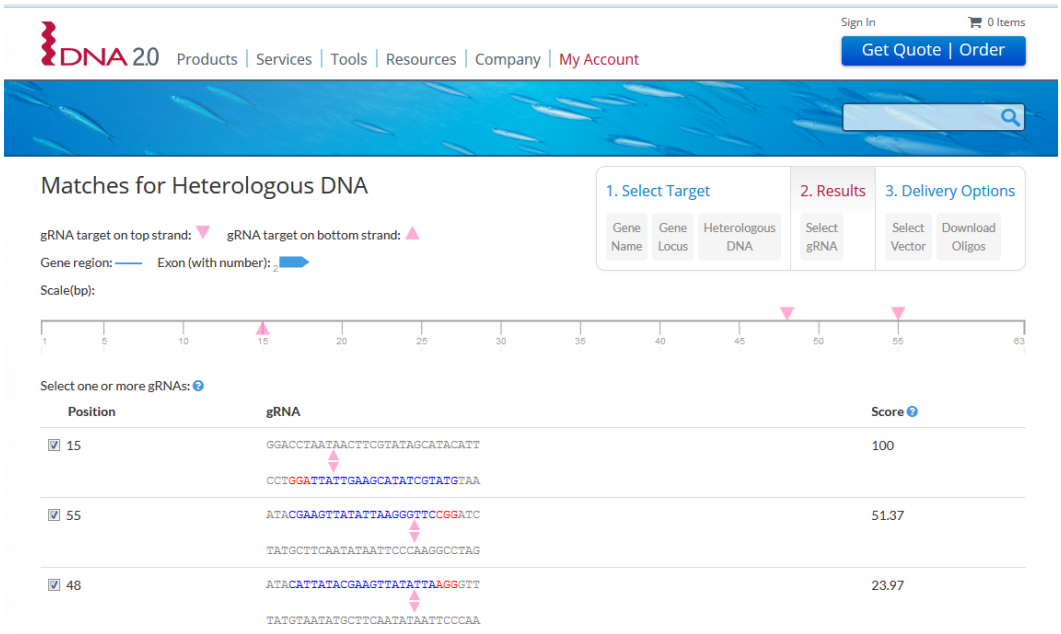


Figure 2.2: Screen capture of the DNA2.0 online software indicating the three potential guides against the loxP region identified. Accessed on 10/02/2015.

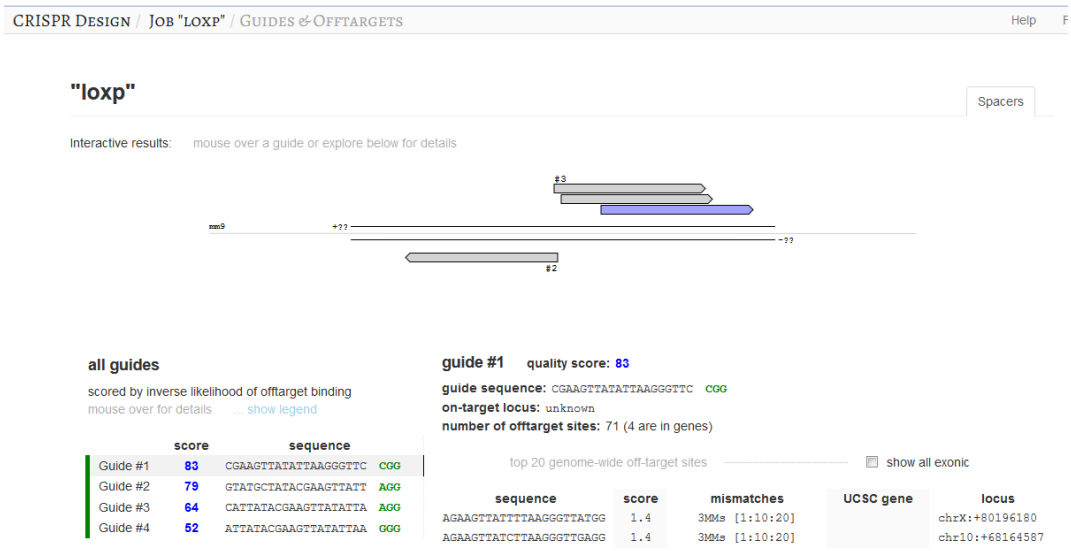


Figure 2.3: Screen capture of the MIT CRISPR Design online software indicating the four potential guides identified against the loxP region. Accessed on 10/06/2015.

Based on this, the following two guides were chosen to target the locus:

Name	Target Site	PAM	MIT Score	DNA2.0 Score	Cut Site Position (nt)
LoxP#1	CGAAGTTATATTAAGGGTTC	CGG	83	51	55
LoxP#2	GTATGCTATACGAAGTTATT	AGG	79	100	15

Table 2.1: Guide-RNAs chosen to target the loxP region in the *mTmG* mice.

These two guides target either side of the loxP site on opposite strands, in case cleavage proceeds more efficiently on one strand than the other:

```
5′ TCGAGGGACCTAATAACTTCGTATAGCATACATTATACGAAGTTATATTAAGGGTTCCGGATC 3′   LoxP#1
3′ AGCTCCCTGGATTATTGAAGCATATCGTATGTAATATGCTTCAATATAATTCCCAAGGCCTAG 5′   LoxP#2
```

Here, the target site is highlighted in yellow with the PAM highlighted in blue.

2.1.2 Zmynd10 Locus

Guides targeting the -7bp deletion locus in the *Zmynd10* mice were chosen using the same procedure as for the *mTmG* guides. A 100bp sequence surrounding the deletion was put into MIT CRISPR Design, DNA2.0 and ZiFit:

```
Zmynd10-001 ENSMUST00000010188

1310 LINE (CRISPR 1)

       661 CTGCAAAAGCAGGCGGAGATGATGGAATTTGAGATATCCCTGAAAGCCCTCTCGGTGCTT
       661 CTGCAAAAGCAGGCGGAGATTTTGAGATATCCCTGAAAGCCCTCTCGGTGCTTCGCTACA
       172 -L--Q--K--Q--A--E--M--M--E--F--E--I--S--L--K--A--L--S--V--L-
       172 -L--Q--K--Q--A--E--I--L--R--Y--P--STOP
      -7bp deletion at 672, missense at aa 178 and PTC aa 185
```

Figure 2.4: DNA and protein sequence of the region surrounding the -7bp deletion in our *Zmynd10* mouse line compared to the reference.

Each program produced several possible guides targeting this region, unfortunately none of the guides were specific to the -7bp deletion mutant and so will also cleave the wild-type allele:

Figure 2.5: Output from the MIT CRISPR Design tool for the *Zmynd10* -7bp deletion locus, showing several potential guides identified. Grey bars are predicted guides relative to the input sequence. No guides specific to the -7bp deletion mutant.

Guide prediction was therefore repeated with a larger input sequence spanning 300bp around the deletion. Again, each program produced several possible guides. From this output three guides were chosen, one on the forward strand and two on the reverse strand, based on the guides' rankings and their proximity to the deletion:

Name	Target Site	PAM	MIT Score	ZiFit Off-target Prediction
Z10 R	GAAGCACCGAGAGGGCTTTC	AGG	73	Summary: Off by 0 = 1; Off by 1 = 0; Off by 2 = 0; Off by 3 = 6
Z10 R2	GATGTAGCGAAGCACCGAGA	GGG	86	Summary: Off by 0 = 1; Off by 1 = 0; Off by 2 = 0; Off by 3 = 5
Z10 F	GATATCCCTGAAAGCCCTCT	CGG	63	Summary: Off by 0 = 1; Off by 1 = 0; Off by 2 = 1; Off by 3 = 10

Table 2.2: Guide-RNAs chosen to target the *Zmynd10* -7bp deletion locus.

Z10 R:

GTAAGCTACGGTTTCCTGGACTTAACAAAATCTCGATGTGAGTTCTCTCCAGCCTCCACAGGCAGGGTGAATGCTAT
GTGCCTGTGCCCCTACCCCCCTCTCCCGCACCATCAGGAGCTGCAAAAGCAGGCGGAGATGATGGAATTTGAGATAT
CCCTGAAAGCCCTCTCGGTGCTTCGCTACATCACAGACTGCGTGGATAGGTGGGCCACCCTCCAAGGCCCGGGGGCG
AGCACACAACAAGGGGCTGTGTGTCCTGGCCCGTGGCCTCTGCTGACCACCCTTCTACCCTACCTCTAG

Z10 R2:

GTAAGCTACGGTTTCCTGGACTTAACAAAATCTCGATGTGAGTTCTCTCCAGCCTCCACAGGCAGGGTGAATGCTAT
GTGCCTGTGCCCCTACCCCCCTCTCCCGCACCATCAGGAGCTGCAAAAGCAGGCGGAGATGATGGAATTTGAGATAT
CCCTGAAAGCCCTCTCGGTGCTTCGCTACATCACAGACTGCGTGGATAGGTGGGCCACCCTCCAAGGCCCGGGGGCG
AGCACACAACAAGGGGCTGTGTGTCCTGGCCCGTGGCCTCTGCTGACCACCCTTCTACCCTACCTCTAG

Z10 F:

GTAAGCTACGGTTTCCTGGACTTAACAAAATCTCGATGTGAGTTCTCTCCAGCCTCCACAGGCAGGGTGAATGCTAT
GTGCCTGTGCCCCTACCCCCCTCTCCCGCACCATCAGGAGCTGCAAAAGCAGGCGGAGATGATGGAATTTGAGATAT
CCCTGAAAGCCCTCTCGGTGCTTCGCTACATCACAGACTGCGTGGATAGGTGGGCCACCCTCCAAGGCCCGGGGGCG
AGCACACAACAAGGGGCTGTGTGTCCTGGCCCGTGGCCTCTGCTGACCACCCTTCTACCCTACCTCTAG

Key: Intron, Exon, -7bp Deletion, Target Site, PAM Site

Figure 2.6: Location of the chosen guides against *Zmynd10* in relation to the deletion.

2.2 Processing of CRISPR gRNAs

Following design, oligonucleotides were ordered from Sigma-Aldrich to allow cloning into several commonly used Cas9/gRNA expression plasmids, as per the protocol in Ran, Hsu et al. (2013). Oligonucleotides were synthesised for both strands following this general template, where N indicates the target site:

5’ - CACC**G**NNNNNNNNNNNNNNNNNNN - 3’

3’ - **C**NNNNNNNNNNNNNNNNNNNCAAA - 5’

Once received, oligonucleotides were resuspended in NF-H_2O (nuclease-free water) to a stock concentration of 100µM. The initial G allows efficient transcription from the U6 promoter, while the added overhangs allow cloning into a BbsI-digested expression vector (Figure 2.7).

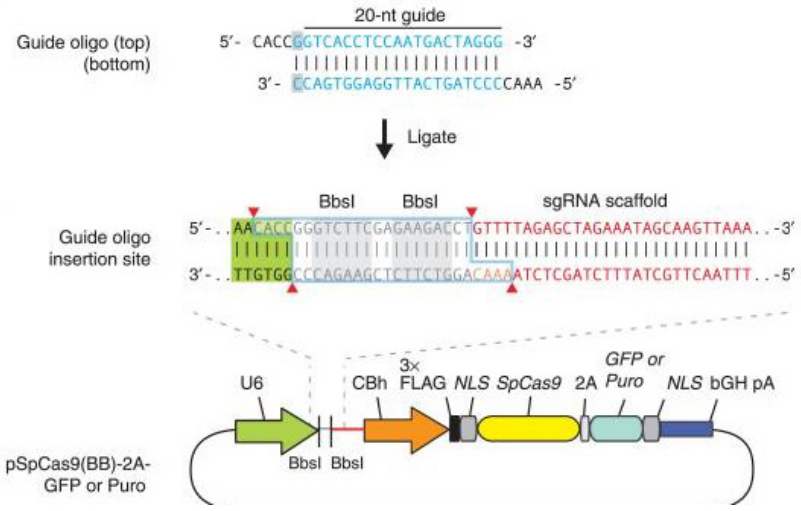


Figure 2.7: Cloning guides into an expression vector. Top and bottom oligonucleotides are first annealed. Subsequently, the added overhangs then allow ligation into a BbsI-digested expression vector, which inserts the guide downstream of the U6 promoter, and immediately upstream of the rest of the sgRNA scaffold. Taken from Ran, Hsu et al. (2013).

2.2.1 Construction of Expression Plasmids

The oligonucleotides were cloned into the expression plasmids, which are listed at the end of this section, following the protocol in 'Genome engineering using the CRISPR-Cas9 system' by Ran, Hsu et al. (2013).

Briefly, the top and bottom oligonucleotides were phosphorylated and annealed using T4 PNK (polynucleotide kinase); incubating at 37°C for 30 minutes, then 95°C for 5 minutes, and then cooling to 25°C at 2.5°C min^{-1} (note: this is half the speed of the protocol in Ran, Hsu et al. (2013) as I found this improved results).

Following this, the annealed oligonucleotides were diluted 1:200 in ddH_2O and then ligated into the BbsI-digested vectors in a reaction with the following components:

Component	**Amount (µl)**
Expression Plasmid (100ng/µl)	1.0
Diluted oligo duplex	2.0
Tango buffer, 10x	2.0
DTT, 10mM	1.0
ATP, 10mM	1.0
FastDigest BbsI	1.0
T4 ligase	0.5
ddH2O	11.5
Total	20.0

This reaction was incubated for a total of 1 hour, in a programme which consisted of six ten-minute cycles of 37°C for 5 minutes, followed by 21°C for 5 minutes, to allow digestion and annealing to take place.

Alternatively, when FastDigest BbsI (Thermo Fisher) was not available, the expression plasmid was first digested with BbsI (NEB) for one hour at 37°C, and then phosphorylated with SAP (shrimp alkaline phosphatase) at 37°C for 30 minutes. Following this, the annealed, diluted oligonucleotides were ligated into the BbsI-digested plasmid with T4 ligase overnight at 16°C.

The ligation product was then treated with PlasmidSafe ATP-dependent exonuclease (Lucigen) at 37°C for 30 minutes to remove any remaining linearized DNA, and then incubated at 70°C for 30 minutes to deactivate any remaining enzymes. Following PlasmidSafe treatment, the plasmid could be stored at -20°C prior to being transformed into DH5α Subcloning Efficiency *Escherichia coli* (*E.* coli) (Thermo Fisher) as per the protocol in Section 2.3.1.

2.2.2 List of Plasmids Used

All Cas9 expression plasmids used were a gift from Feng Zhang (Ran, Hsu et al. 2013), and are available on Addgene (https://www.addgene.org). The plasmid stocks used to perform the cloning reactions in Section 2.2.1 were prepared and sequence- verified by Peter Budd.

2.2.2.1 pX330 (Addgene plasmid #42230)

The 'basic' sgRNA, Cas9 expression plasmid which contains no selection cassette:

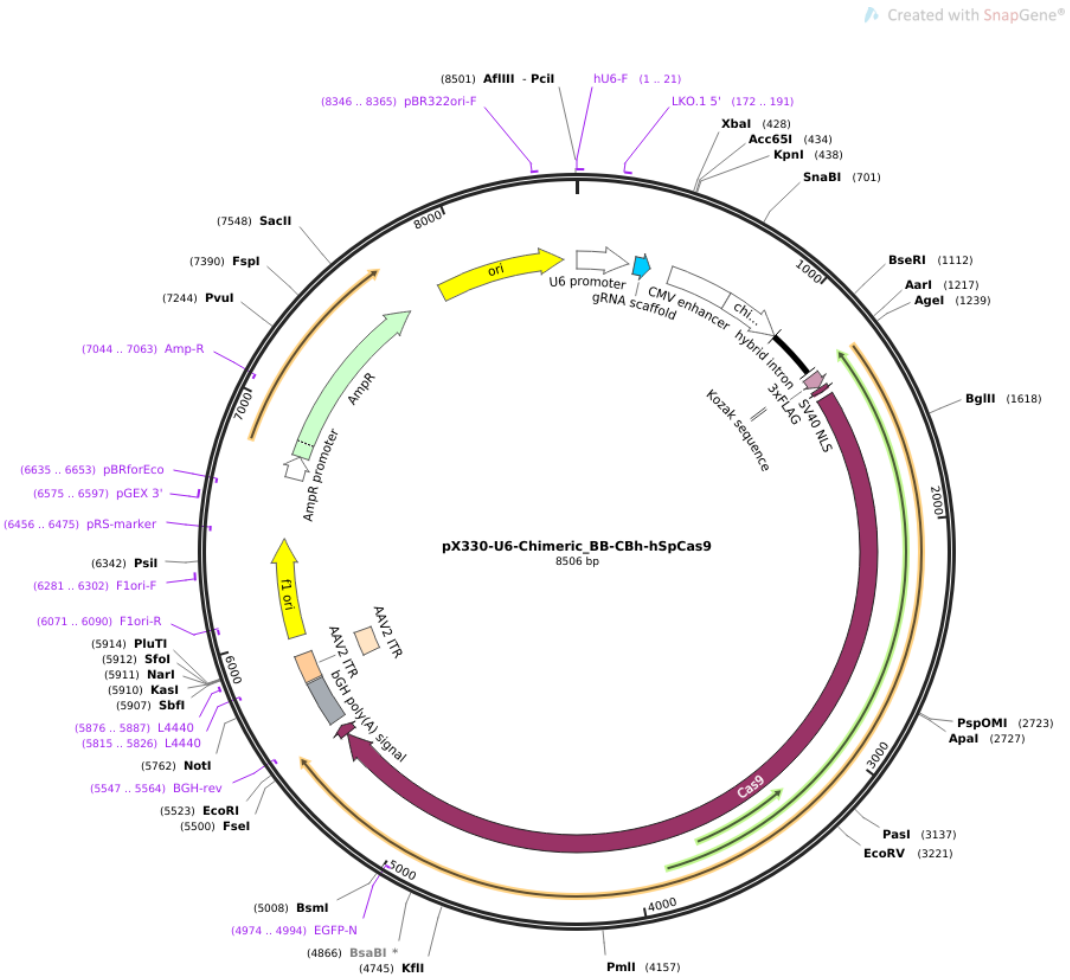


Figure 2.8: Diagrammatic representation of the pX330 expression plasmid. Taken from Addgene, https://www.addgene.org/42230/, accessed on 03/01/2018. A gift from Feng Zhang (Ran, Hsu et al. 2013).

2.2.2.2 pX458 (Addgene plasmid #48138)

A derivative of pX330 which contains an EGFP reporter following Cas9:

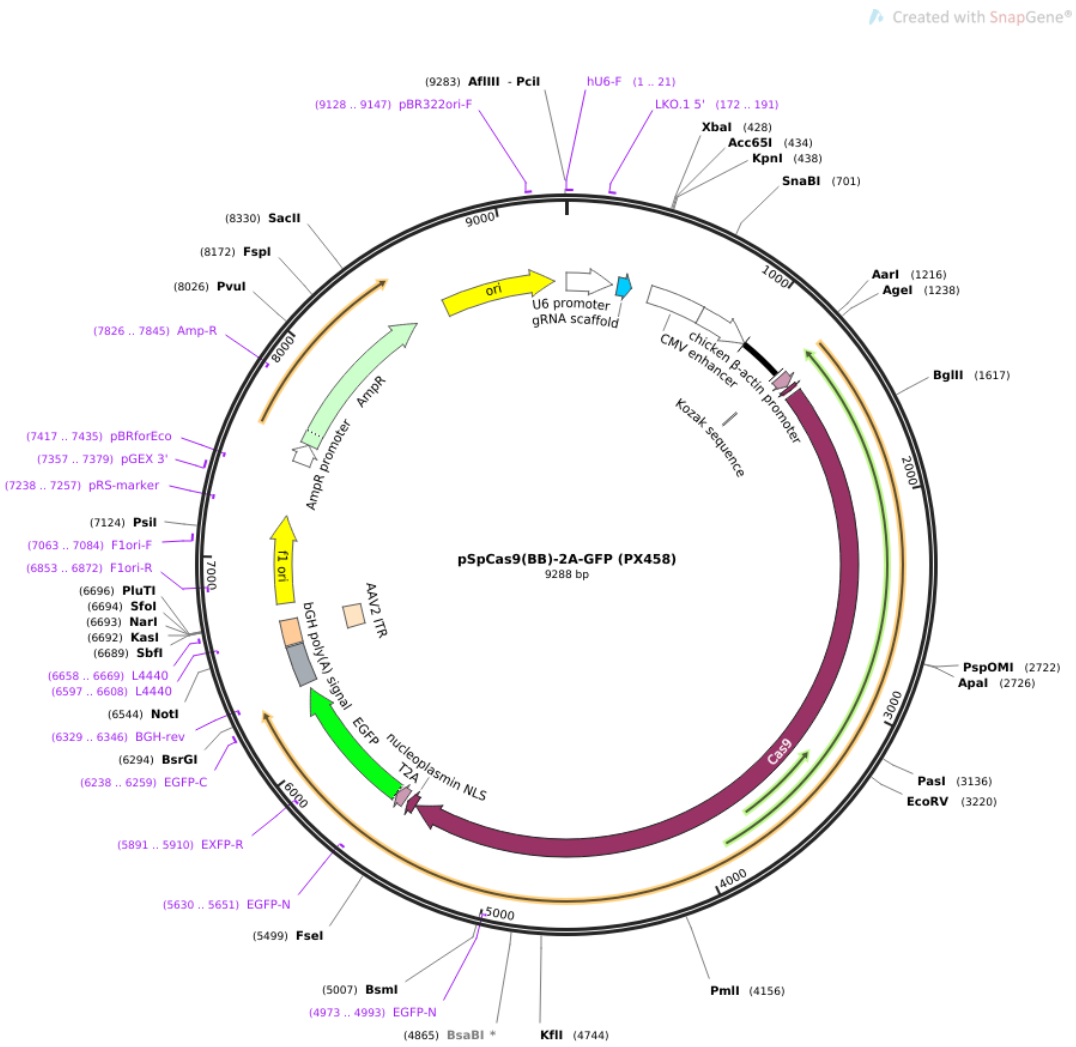


Figure 2.9: Diagrammatic representation of the pX458 expression plasmid. Taken from Addgene, https://www.addgene.org/48138/, accessed on 03/01/2018. A gift from Feng Zhang (Ran, Hsu et al. 2013).

2.2.2.3 pX459 V2.0 (Addgene plasmid #62988)

A derivative of pX330 which contains a puromycin resistance cassette following Cas9:

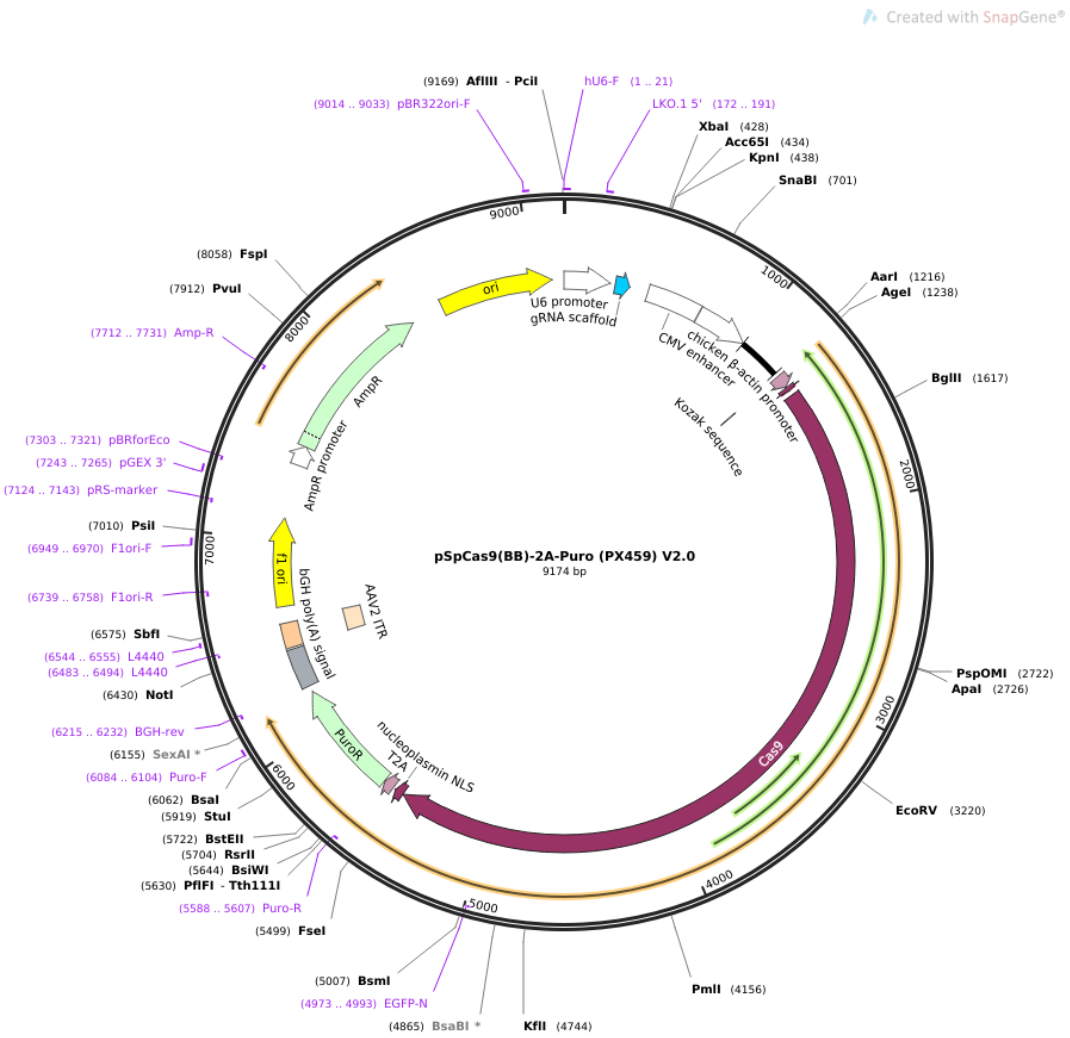


Figure 2.10: Diagrammatic representation of the pX459 v2.0 expression plasmid. Taken from Addgene, https://www.addgene.org/62988/, accessed on 03/01/2018. A gift from Feng Zhang (Ran, Hsu et al. 2013).

As the original version of this plasmid was not used, the name 'pX459' in this thesis refers exclusively to V2.0.

2.2.2.4 pmaxGFP (Lonza)

The GFP expression vector pmaxGFP was included in the P4 Primary Cell 4D-Nucleofector X Kit L (Lonza, catalogue number: V4XP-4024), and was used for several optimisation experiments.

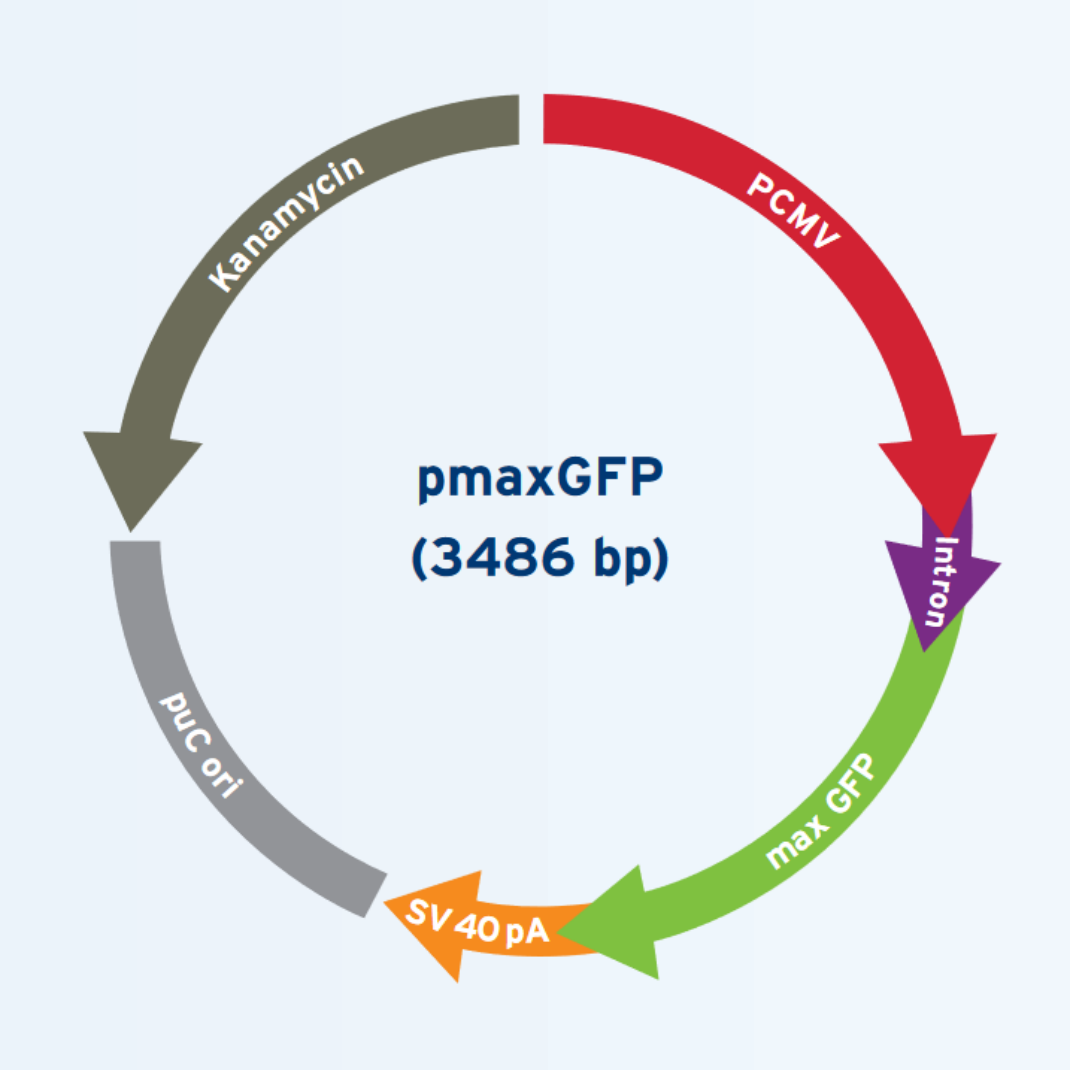
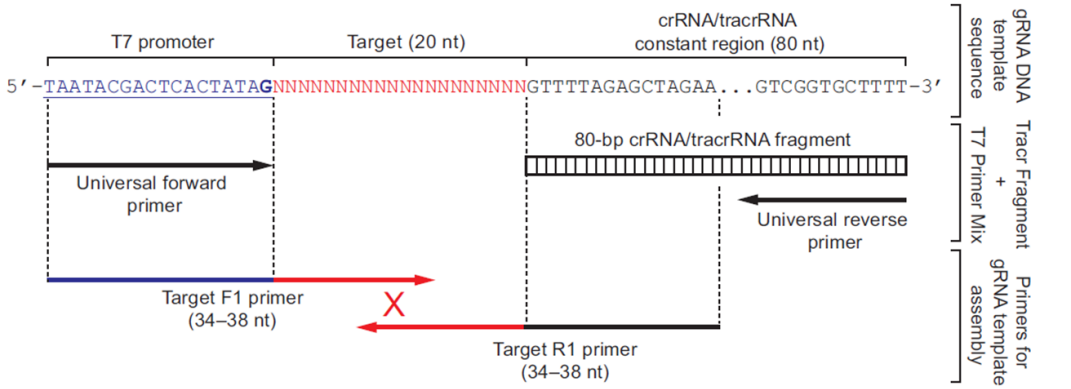


Figure 2.11: Diagrammatic representation of the pmaxGFP vector. Taken from http://bio.lonza.com/fileadmin/groups/FAQs/public/Technology_Flyer.pdf, accessed 29/01/2018.

2.2.3 Preparation of sgRNAs for Ribonucleoprotein (RNP) Delivery

As an alternative approach to a plasmid-based expression system, direct delivery of the Cas9 protein complexed to the sgRNA was also attempted. The design and synthesis of these complexes was based on the user guide supplied with the GeneArt Platinum Cas9 Nuclease (catalogue number B25641) from Thermo Fisher Scientific (2015).

The initial gRNA DNA template was generated using the GeneArt Precision gRNA Synthesis Kit, also from Thermo Fisher (catalogue number A29377), based on the manufacturer's instructions (Thermo Fisher Scientific 2016). Primers were designed to insert the target sequence between the T7 promoter and crRNA/tracrRNA as so:



Target F1: TAATACGACTCACTATAG + first 16nt of the target sequence
Target R1: TTCTAGCTCTAAAAC + first 19nt of the target sequence reverse complement

Figure 2.12: Schematic of primer design for the production of the gRNA DNA template sequence, for later *in vitro* transcription to produce a sgRNA. Adapted from Thermo Fisher Scientific (2016).

The primers were then ordered from Sigma-Aldrich. Once received they were resuspended in NF-H_2O to make a stock solution at 10µM, and a working solution at 0.3µM. Following this, a PCR assembly reaction was set up as so:

Phusion High-Fidelity PCR Master Mix (2x) (NEB, cat. no. M0531S)	12.5µl
Tracr Fragment + T7 Primer Mix	1.00µl
0.3µM Target F1/R1 oligonucleotide mix	1.00µl
NF-H_2O	10.5µl
Total	25.00µl

A PCR was then performed to amplify the gRNA DNA template with the following cycling parameters:

Temperature	Time	Number of Cycles
98°C	10s	1x
98°C 55°C	5s 15s	32x
72°C	1min	1x
4°C	hold	1x

After this, an *in vitro* transcription (IVT) reaction was set up with the following components and incubated at 37°C for 2 hours:

NTP Mix (100mM of each dNTP)	8µl
gRNA DNA Template	6µl
5X TranscriptAid Reaction Buffer	4µl
TrascriptAid Enzyme Mix	2µl
Total	20µl

Once the IVT reaction was complete, 1µl of DNase I was added to the reaction mix and it was again incubated at 37°C for 15 minutes. The *in vitro* transcribed gRNA was then purified as per the manufacturer's instructions using the kit components and eluted in NF-H_2O. Purified sgRNAs were then stored at -80°C before being analysed.

2.2.3.1 RNA Purity Analysis

Following purification, all sgRNAs were denatured at 70°C for 2 minutes and analysed on an Agilent 2100 Bionalayzer, using a nanogram sensitivity eukaryote total RNA assay, to check their purity and the size and concentration of the product. If a single product of the correct size was present, the sgRNA was diluted to 240ng/µl in NF-H_2O and aliquoted for future use. All aliquots were stored at -80°C.

2.3 Bacterial Work

Once guides had been cloned into the sgRNA/Cas9 expression plasmids as detailed in Section 2.2.1, the resulting PlasmidSafe-treated ligation mixes were then transformed into *E. coli*.

2.3.1 Transformation of Plasmids

Subcloning Efficiency DH5α Competent *E. coli* (Thermo Fisher, catalogue number: 18265017) were used for the transformation of all the plasmids in this thesis. Transformations were performed as per the manufacturer's instructions, adding 2µl of the PlasmidSafe-treated products from 2.2.1 to each 50µl aliquot of *E. coli*. Following transformation, several different volumes were plated on ampicillin-containing LB-plates (Luria Bertani-plates) to ensure there were well-spaced colonies, and plates were incubated overnight at 37°C. The next day, if transformation had been successful, plates were stored at 4°C if they were not immediately used in the following steps.

2.3.2 Isolation of Plasmids

2.3.2.1 Colony Re-spreading

If transformation was successful, and several well-spaced colonies were found the next day, two or three of these colonies were then re-spread on new ampicillin LB-plates. This was done to reduce the risk of accidentally picking up two overlapping colonies and to remove any non-antibiotic resistant bacteria which may be on the plate.

To do this, a single colony was stabbed with a pipette tip, which was then inserted into a 1.5ml microcentrifuge tube with 1ml LB. Using a Gilson pipette, the contents of the tip were then aspirated in and out several times to disperse the bacteria into the LB. Following this, 1µl was taken from the tube and added to another microcentrifuge tube with 1ml LB to create a 1 in 1000 dilution, this was then vortexed briefly. 10µl of this 1 in 1000 dilution was then spread out on an ampicillin-containing LB-plate. Generally, this would give 10-50 well-spaced colonies the next day. If successful, the resulting plates were then stored at 4°C if they were not immediately used for mini-preps.

2.3.2.2 Mini-preps

Mini-preps were performed using a QIAprep Spin Miniprep Kit (QIAGEN, catalogue number: 27106) as per the manufacturer's instructions. A re-spread colony from the preceding step was inoculated into 3ml of LB + ampicillin in a 15ml Falcon tube, which was then placed at 37°C overnight in a shaking incubator. The next morning, 2ml of the bacterial cultures was harvested by centrifugation and the manufacturer's

protocol was followed to extract total DNA, eluting in either NF-H_2O or TE buffer. The remaining culture was stored at 4°C for potential future use. The purity and concentration of the eluted DNA was then checked using a Nanodrop spectrophotometer (Section 2.4.3). If satisfactory, the plasmids were sent for sequencing using the U6 Forward primer, which will reveal if the guide sequence has been correctly inserted (Section 2.9):

U6 Forward Primer GAGGGCCTATTTCCCATGATTCC

2.3.2.3 Maxi-preps

If sequencing indicated the correct insertion of the guides into the expression plasmids, a maxi-prep was then performed. Firstly, 100µl of the stored outgrowth from the mini-prep was inoculated into 100ml of LB + ampicillin in a 1 litre conical flask. This was then incubated with vigorous shaking overnight at 37°C. Following this, the bacterial culture was harvested by centrifugation and the plasmids were extracted using a QIAGEN Plasmid Maxi Kit (QIAGEN, catalogue number: 12163), following the manufacturer's instructions. In the final step, plasmids were eluted in 250µl of TE buffer and the concentration and purity of the preparation was checked on a Nanodrop spectrophotometer (Section 2.4.3). All plasmids were then stored at 4°C for future use.

2.4 DNA Manipulation

2.4.1 DNA Extraction

2.4.1.1 DNA Extraction from Cells

DNA was extracted from cells, generally in a 6-well plate, using Quick Lysis Buffer (QLB, Section 2.22.3). Firstly, the media was aspirated from the wells and the cells were rinsed in PBS. Following this, 250µl QLB + 10µl Proteinase K was added to each of the wells (quantities were scaled up or down for different well sizes). The plates were then placed at 55°C for between 2 and 16 hours. If the QLB had then evaporated, the contents of the well were resuspended in extra QLB + Proteinase K and transferred to a 1.5ml microcentrifuge tube. This was then centrifuged at maximum speed (>14,000 x g) at room temperature for 5 minutes. Following this, the supernatant was transferred to a new tube and 250µl isopropanol was added. The tube was then vortexed briefly and incubated at room temperature for 30 minutes. Subsequently, the

tube was centrifuged at maximum speed at room temperature for 30 minutes. The supernatant was removed from the tube and the DNA pellet was washed in 70% ethanol. Excess ethanol was carefully removed from the tube and the pellet was then left to dry at room temperature for approximately 20 minutes. The DNA pellet was then resuspended in 100µl TE buffer and the concentration and purity of the DNA extraction was checked on a Nanodrop spectrophotometer (Section 2.4.3).

2.4.1.2 DNA Extraction from Mouse Ear Punches and Tail Tips

Ear punches were taken from mice at weaning for genotyping, or alternatively tail tips during dissections, and collected in a microcentrifuge tube. To quickly extract DNA from these tissue samples, 100µl of buffer 25mM NaOH/0.2mM EDTA was added to each tube and they were then incubated at 95°C for 20 minutes. Following this, 100µl of 40mM Tris-HCl was added to each tube, which was then vortexed. The extracted DNA could then be checked on a Nanodrop spectrophotometer (Section 2.4.3), although generally 2µl was used in the following genotyping PCR (Section 2.6).

2.4.1.3 DNA Extraction from Gel Slices

To isolate a single band of interest from an agarose gel (Section 2.7), generally after a PCR reaction, the gel was first visualised under a UV transilluminator. The desired band was then quickly and accurately excised using a clean scalpel and placed into a microcentrifuge tube. Following this, DNA was extracted from the gel slice using a QIAquick Gel Extraction Kit (QIAGEN, catalogue number: 28706), as per the manufacturer's instructions. In the final step, DNA was eluted in 30µl of TE buffer and then either checked on a Nanodrop spectrophotometer (Section 2.4.3) or visually verified on another agarose gel (Section 2.7).

2.4.2 DNA Purification

To isolate DNA and remove contaminants such as enzymes and dNTPs from a PCR reaction, or to change DNA between buffers, a DNA purification was performed. All DNA purifications were carried out using a QIAquick PCR Purification Kit (QIAGEN, catalogue number: 28106) as per the manufacturer's instructions. In the final step, DNA was eluted in an appropriate volume of TE buffer, or alternatively NF-H_2O.

2.4.3 DNA Quantification

To check concentration and purity, DNA samples were analysed on a NanoDrop1000 spectrophotometer (Thermo Fisher). Concentration was gauged by the absorbance at 260nm, while purity was gauged by the ratio of absorbances at 260nm/280nm. A sample was considered pure if it had an A260/A280 of 1.8 or greater.

2.4.4 DNA Digestion

2.4.4.1 Restriction Enzyme Digests

The only restriction enzyme digest performed in this thesis used MlucI, purchased from New England Biolabs (catalogue number: R0538S), for genotyping the *Zmynd10* -7bp deletion mutants. To perform the restriction enzyme digest, 15µl of the genotyping PCR product (Section 2.5.2) was taken. To the PCR product, 0.5µl of MlucI was added, along with 3µl of 10x CutSmart buffer (NEB) and NF-H_2O up to a total volume of 30µl. The PCR product was digested at 37°C for 30 minutes, and the digestion products were then analysed on a 2% agarose gel (Section 2.7).

2.4.4.2 Surveyor Assay

Surveyor assays were carried out using the Surveyor Mutation Detection Kit (IDT DNA, catalogue number: 706020), as per the manufacturer's instructions. Prior to this, PCR products were purified (Section 2.4.2) and resuspended in Surveyor buffer + NF-H_2O (Section 2.22.1) for optimal results.

2.5 Primer Design

All primers were designed using the Primer3 online tool (Koressaar and Remm 2007, Untergasser, Cutcutache et al. 2012), available at http://primer3.ut.ee/, using the relevant mis-priming library, i.e. rodent. Furthermore, all primers were checked for specificity in the mouse genome using the UCSC *in silico* PCR (http://rohsdb.cmb.usc.edu/GBshape/cgi-bin/hgPcr), and in the sequence of interest using BLAST (https://www.ncbi.nlm.nih.gov/BLAST/) (where the sequence would not be present in the mouse genome).

2.5.1 *mTmG* Sequencing Primers

Six primers were designed to sequence across the *mTmG* locus:

Primer 1	-	ACGTGCTGGTTATTGTGCTG
Primer 2	-	TACCTTCACGTGGCCATTCT
Primer 3	-	CTTGGGCTGCAGGTCGAG
Primer 4	-	GTCTTGTAGTTGCCGTCGTC
Primer 5	-	CCATGTTGTTGTCCTCGGAG
Primer 6	-	TGATGAATGGGAGCAGTGGT

These primers were used in the following combinations:

Primer 1 + Primer 5	**-**	914bp
Primer 6 + Primer 2	**-**	917bp
Primer 3 + Primer 4	**-**	538bp
Primer 1 + Primer 4	-	3017bp

Further details of the sequencing experiment are in Chapter 4.

2.5.2 *mTmG* Promoter Sequencing Primers

Two sets of primers, three primers in total, were designed for sequencing across the promoter region; one forward primer and two reverse primers, one in tdTomato and one in EGFP.

Tomato primers (927bp):

***mTmG*_Tomato_Promoter_Seq_Left**	-	CCTCCCCGAGTTGCTGAG
***mTmG*_Tomato_Promoter_Seq_Right**	-	CTTGGAGCCGTACATGAAC

EGFP primers (931bp):

***mTmG*_EGFP_Promoter_Seq_Left**	-	CCTCCCCGAGTTGCTGAG
***mTmG*_EGFP_Promoter_Seq_Right**	-	GGTGCAGATGAACTTCAGGG

2.5.3 *Zmynd10* Genotyping Primers

Two primers were used for the genotyping of *Zmynd10* -7bp deletion mutants, these were both designed by Dr. Girish Mali.

766 - CCAGGACAGTACCCCTATGC

767 - AGGTAGGGTAGAAGGGTGGT

2.5.4 *Zmynd10* Ion Torrent Primers

A series of barcoded forward primers were designed to amplify the *Zmynd10* samples for Ion Torrent (Thermo Fisher) sequencing, based on primers 766 and 767:

Primer	Sequence	Sample Name
Pr 766_1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGATCC AGGACAGTACCCCTATGC	**B6**
Pr 766_2	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGATC CAGGACAGTACCCCTATGC	**1Bl**
Pr 766_3	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGATC CAGGACAGTACCCCTATGC	**2Bl**
Pr 766_4	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGATCC AGGACAGTACCCCTATGC	**3Bl**
Pr 766_5	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGATC CAGGACAGTACCCCTATGC	**4Bl**
Pr 766_6	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCGATCC AGGACAGTACCCCTATGC	**5Bl**
Pr 766_7	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGATCC AGGACAGTACCCCTATGC	**6Bl**
Pr 766_8	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACGATCC AGGACAGTACCCCTATGC	**7Bl**
Pr 766_9	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGCGGAACGATC CAGGACAGTACCCCTATGC	**8Bl**
Pr 766_10	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACCGAACGATCC AGGACAGTACCCCTATGC	**2Gr**
Pr 766_11	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTCGAATCGATCC AGGACAGTACCCCTATGC	**3Gr**
Pr 766_12	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGGTGGTTCGATCC AGGACAGTACCCCTATGC	**4Gr**
Pr 766_13	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAACGGACGATCC AGGACAGTACCCCTATGC	**5Gr**
Pr 766_14	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGTGTCGATCC AGGACAGTACCCCTATGC	**6Gr**
Pr 766_15	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGAGGTCGATCC AGGACAGTACCCCTATGC	**7Gr**
Pr 766_16	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGATGACGATCC AGGACAGTACCCCTATGC	**8Gr**

Table 2.3: *Zmynd10* Ion Torrent sequencing primers.

These use IonXpress barcodes and adapter sequences (Thermo Fisher). All samples were amplified with the same reverse primer:

Pr 767_trP1 -

CCTCTCTATGGGCAGTCGGTGATAGGTAGGGTAGAAGGGTGGT

2.5.5 *Zmynd10* Sequencing Primers

Four primers were designed to sequence across a larger region surrounding the *Zmynd10* -7bp deletion locus in our mouse lines, to check for variants:

Set 1 (521bp):

Z10_SEQ_1	**-**	CTGTCTCTCCCTCTCAGCC
Z10_SEQ_2	**-**	GCAGTCTGTGATGTAGCGAA

Set 2 (589bp):

Z10_SEQ_3	**-**	AGCTACGGTTTCCTGGACTT
Z10_SEQ_4	**-**	ACGTCAGAACAAGGTGGAGT

2.6 Polymerase Chain Reaction

The polymerase chain reaction (PCR) was used to amplify DNA, using the primers mentioned in the previous section (Section 2.5). All reactions were either performed with Taq polymerase, DreamTaq Green PCR Master Mix (Thermo Fisher, catalogue number: K1081), or a proof-reading polymerase, Phusion High-Fidelity DNA Polymerase (Thermo Fisher, catalogue number: F530S). Generally, Taq polymerase was used for simple genotyping experiments, whereas Phusion polymerase was used for next-generation sequencing experiments.

2.6.1 PCR Reaction Compositions

For *Zmynd10* genotyping, **DreamTaq Green PCR Master Mix** was used:

DreamTaq Green PCR Master Mix	15.0µl
NF-H_2O	11.5µl
Forward Primer (10µM)	0.75µl
Reverse Primer (10µM)	0.75µl
Template	2.00µl
Total	30.0µl

For *Zmynd10* Ion Torrent (Thermo Fisher) sequencing, **Phusion HF Master Mix** (Thermo Fisher, catalogue number: M0531S) was used:

2xPhusion HF Master Mix	10.0µl
NF-H_2O	5.40µl
Forward Primer (10µM)	1.00µl
Reverse Primer (10µM)	1.00µl
DMSO	0.60µl
Template	2.00µl
Total	20.0µl

For *mTmG* sequencing experiments, **Phusion Hot Start II DNA Polymerase** (Thermo Fisher, catalogue number: F549L) was used. This was used in combination with various concentrations of betaine, for the highest specificity.

For primer sets 1 + 5, 6 + 2 and 3 + 4 the following composition was used:

NF-H2O	24.5µl
Phusion HF Buffer	10.0µl
Betaine (5.0M)	9.00µl
10mM dNTPs	1.00µl
Phusion Hot Start II DNA Polymerase	0.50µl
Forward Primer (10µM)	2.00µl
Reverse Primer (10µM)	2.00µl
Template DNA	1.00µl
Total	50.0µl

Whereas for primer set 1 + 4, as well as the promoter sequencing primers, the following composition was used:

NF-H2O	18.5µl
Betaine (5.0M)	15.0µl
Phusion GC Buffer	10.0µl
10mM dNTPs	1.00µl
Phusion Hot Start II DNA Polymerase	0.50µl
Forward Primer (10µM)	2.00µl
Reverse Primer (10µM)	2.00µl
Template DNA	1.00µl
Total	50.0µl

2.6.2 Cycling Parameters

Cycling parameters were optimised for each PCR reaction, and as such annealing temperatures and extension times were adjusted to suit each set of primers. Furthermore, cycle numbers were optimised for different samples to achieve optimal yields with no non-specific bands. However, general cycling parameters for each of the DNA polymerases is included below.

DreamTaq Green PCR Master Mix (*Zmynd10* genotyping):

Step	Temperature (°C)	Time	Number of Cycles
Initial Denaturation	95	1-3 min	1
Denaturation	95	30s	25-40
Annealing	Tm - 5	30s	
Extension	72	1min/Kb	
Final Extension	72	5-15 min	1

Phusion HF Master Mix (*Zmynd10* Ion Torrent sequencing):

Step	Temperature (°C)	Time	Number of Cycles
Initial Denaturation	98	1 min	1
Denaturation	98	10s	40
Annealing	61.5	22s	
Extension	72	30s	
Final Extension	72	5 mins	1

Phusion Hot Start II (*mTmG* sequencing):

Step	Temperature (°C)	Time	Number of Cycles
Initial Denaturation	98	1 min	1
Denaturation	98	10s	10
Annealing	75 – 65 / 74 - 64	15s	
Extension	72	30s/Kb	
Denaturation	98	10s	29-32
Annealing	65 / 64	15s	
Extension	72	30s/Kb	
Final Extension	72	5 mins	1

In relation to the *mTmG* sequencing primers, for primer set 1 + 4 an annealing temperature of 65°C was used along with an extension time of 1 minute 35 seconds, in a total of 34 cycles. For the other three primer sets an annealing temperature of 64°C was used along with an extension time of 25 seconds, in a total of 39 or 42 cycles.

2.7 Agarose Gel Electrophoresis

DNA was separated by size using agarose gel electrophoresis. Agarose gels were made to a concentration of 1.5% or 2% using Hi-Pure Low EEO agarose (BioGene, catalogue number 300-300) dissolved in TBE buffer. To visualise the DNA molecules in the gel, ethidium bromide (EtBr) was added to the agarose at a concentration of 0.5µg/ml. Once prepared, the melted agarose was poured into gel trays with a comb to form wells. DNA was mixed with Orange G Loading Dye (Section 2.22.1) prior to loading, unless the PCR master mix already contained a loading dye. To mark the sizes of the migrating bands, 5µl of 1 Kb Plus DNA Ladder (Thermo Fisher, catalogue number: 10787018) was added to the side wells. A voltage of 50-150V was then applied to the gel in an electrophoresis gel tank to mobilise the nucleic acids. After one to two hours, depending on the size of the expected DNA fragment(s), the gel was visualised and photographed using a UV transilluminator (BioDoc-It Imaging System, UVP).

2.8 TOPO-Cloning

TOPO-Cloning was performed using the TOPO TA Cloning Kit for Sequencing, and One Shot TOP10 Chemically Competent E. coli (Thermo Fisher, catalogue number: K457540) per the manufacturer's instructions (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/topotaseq_man.pdf).

LB-ampicillin agar plates were spread with X-Gal (20mg/ml) and 100mM IPTG (to allow screening of colonies) and left to dry for 1 hour prior to plating the transformed bacteria. The following day, white or light blue colonies, which should harbour the correct insert, were taken for outgrowth and sequencing. Colonies were stabbed with a pipette tip and added to 1ml of LB + ampicillin in a 96-well plate. These plates were left overnight in a 37°C incubator with vigorous shaking. The following day, the plates were submitted to the IGMM technical services who performed plasmid extractions and sequencing (Section 2.9) with the M13 Forward and M13 Reverse primers:

M13 Forward - GTAAAACGACGGCCAG

M13 Reverse - CAGGAAACAGCTATGAC

2.9 Sanger Sequencing & Analysis

All DNA sequencing was performed by the IGMM technical services department on an Applied Biosystems 3130 (4-capillary) Genetic Analyzer or a 48-capillary 3730 DNA Analyzer (Both Thermo Fisher). When received, sequencing traces were visually inspected using the FinchTV 1.4 software (Geospiza Inc.) and further analysed if necessary using alignment tools such as SnapGene (GSL Biotech LLC) and BLAST (NCBI NLM, https://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.10 Mouse Embryonic Fibroblast Work

2.10.1 Generation of Mouse Embryonic Fibroblasts

Mouse embryonic fibroblasts (MEFs) were generated from mice embryos between the stages of embryonic day (E) 11.5 to E13.5. Briefly, gestational sacs were removed from the uterus and placed into PBS. The embryos were then dissected out of their gestational sacs, and the red organs and heads of the embryos were further removed. A small tissue sample was generally taken for genotyping and placed in QLB for DNA

Extraction (Section 2.4.1.2). Each embryo was then placed into a single well of a 6-well plate, containing 2ml Trypsin/Versene, and cut using forceps into several small pieces. Plates were left at room temperature in a tissue culture hood for an hour to allow cells to dissociate. Following this, 3ml of MEF growth media was added to each well and the contents were transferred to a 15ml Falcon tube. Falcon tubes were centrifuged at 150 x g for 5 minutes to pellet the cells. Pellets were then resuspended in 10ml of MEF growth media and transferred to a T75 flask. The following day, any debris was removed by changing the media. At 90% confluency cells were split 1 in 3. MEF cells were considered ready to use once they had reached passage number 3, as at this point other cell types will have detached.

2.10.2 Culture Conditions

MEFs were maintained in a 37°C, 5% CO_2, humidified, normoxic incubator in MEF growth media. At 90% confluency non-immortalised MEFs were rinsed in PBS and split with Trypsin/Versene (Section 2.22.3) 1 in 3 into a new flask. Immortalised MEFs were split 1 in 10.

2.10.3 Mycoplasma Testing & Liquid Nitrogen Storage

All MEF cultures were tested for mycoplasma at P3 and prior to beginning experiments. Furthermore, all cultures were tested prior to being cryostored. All cultures used were mycoplasma negative.

Generally, non-immortalised MEF cultures were initially expanded and then cryostored once they had reached confluency at P2 stage. Immortalised MEF lines were cryostored immediately following the immortalisation protocol (Section 2.10.4), or sporadically as necessary. To do this, flasks were rinsed in PBS and cells were detached with Trypsin/Versene (Section 2.22.3). Detached cells were centrifuged at 150 x g for 5 minutes and pellets were resuspended in 1ml MEF freezing media (Section 2.22.3) in a cryotube. Cryotubes were immediately put on dry ice and then stored at -80°C prior to liquid nitrogen storage.

2.10.4 Immortalisation Protocol

MEFs were transfected at an early stage, i.e. P3, with an immortalisation plasmid containing an SV40 cassette and a puromycin resistance cassette. This immortalisation

plasmid, pJKN114, was a gift from the lab of Andrew Jackson. MEFs were transfected using the Neon system (Section 2.12.1). A non-transfected control well was also prepared. Twenty-four hours later, media was replaced with MEF growth media + 3µg/ml puromycin. Wells were left until all cells in the control well had detached, generally 3 or 4 days. The puromycin-containing media was removed and wells were rinsed in PBS. Antibiotic-free MEF growth media (Section 2.22.3) was added and wells were allowed to recover for several days. Upon recovery, cells were tested for mycoplasma and either frozen down (Section 2.10.3) or expanded for use in future experiments.

2.11 Mouse Ependymal Cell Work

2.11.1 Generation of Mouse Ependymal Cells

To generate ependymal cells, mice at E18.5 were killed by decapitation and their heads were received. Brains were then manually dissected out into Hank's solution (Section 2.22.3), and from brains the ventricular cups were removed. Ventricular cups, separated by genotype and/or mouse, were placed into appropriately sized wells and manually cut into pieces. TrypLE Express (Thermo Fisher, catalogue number: 12605010) was added to digest the pieces and plate(s) were incubated at 37°C, 5% CO_2 for 60 minutes. Digestion was stopped with the addition of an equal amount of ependymal growth medium (Section 2.22.3). Cell suspensions were then centrifuged for 5 minutes at 150 x g. Cell pellets were resuspended in ependymal growth medium and placed in poly-L-lysine coated wells or flasks. Ependymal progenitor cells were allowed to proliferate until confluence, generally ~5 days, during which time debris and detached cells were regularly removed by washing in 1x PBS and replacing the ependymal growth media.

2.11.2 Culture Conditions & Differentiation

Ependymal progenitor cells and ependymal cells were both maintained in a humidified, normoxic, 5% CO_2, 37°C incubator. Upon confluency, progenitor cells were washed in 1x PBS, and serum-starved in ependymal serum starvation media (1% FCS, Section 2.22.3). This media was regularly replaced until the cells developed multicilia.

2.12 Primary Cell Transfection & Transduction

2.12.1 Neon System

The Neon Transfection System (Thermo Fisher, catalogue number: MPK5000) was employed for both MEF and ependymal cell transfections, using both 10µl and 100µl electroporation kits (catalogue numbers MPK1096 and MPK10096 respectively), and identical electroporation parameters. Mixes for transfection were scaled up or down as necessary between the two tip sizes.

2.12.1.1 Protocol for Neon Plasmid Transfection (MEFs)

Prior to transfection, MEF growth media was added to the wells of a plastic- or glass-bottomed plate as necessary, and small molecules (Section 2.14) were added if required. Plates were then left at 37°C for later use. Buffer R and buffer E2 (100µl electroporation kit, Thermo Fisher) were allowed to warm to room temperature.

MEFs were rinsed in PBS and detached with Trypsin/Versene. An equal volume of MEF growth media was then added to the flask(s) and the contents were transferred to a 15ml Falcon tube, which was centrifuged at 150 x g for 5 minutes. Cell pellets were resuspended in 10ml PBS, and 100µl of this suspension was then diluted 1 in 2 in 100µl of PBS. Using this final dilution, a cell count was taken using a Scepter 2.0 Cell Counter (Merck). Generally, $0.5x10^5$ cells were used per 10µl transfection, or $5x10^5$ cells per 100µl transfection, and an excess was always prepared. Based on the cell count, an appropriate volume of cells in PBS was centrifuged at 150 x g for 5 minutes.

During centrifugation, transfection mixes were prepared in microcentrifuge tubes. In all experiments, 1000ng of Cas9 expression plasmid and 1000ng of repair plasmid/minicircle (if included) were used per $1x10^5$ cells (up to a total of 2000ng), unless otherwise noted. If an ssODN repair template was included, 1µl of ssODN at 10µM concentration was used per $2x10^5$ cells. Following centrifugation, the supernatant was removed and the pellet was resuspended in an appropriate volume of Buffer R to bring the final volume to 10µl or 100µl per transfection, taking account of the volume of plasmid(s) or ssODN used. Cells were then added to the transfection mixes.

For transfection, a Neon tube (Thermo Fisher) was placed in the Neon Transfection System and 3ml of Buffer E2 was added. The cell-DNA mixture was taken up into the Neon electroporation tip and electroporated with the following parameters: voltage 1350V, width 30ms and pulses 1, unless otherwise noted. Following electroporation, cells were immediately returned to MEF growth media (Section 2.22.3). Neon tips and Neon tubes were changed between samples, however, when transfecting the same sample they were reused up to 5 times. Following electroporation, plates were incubated at 37°C, 5% CO_2 overnight and the following day the media was replaced.

2.12.1.2 Protocol for Neon Plasmid Transfection (Ependymal Cells)

Prior to transfection, ependymal growth media was added to the wells of laminin-coated plastic- or glass-bottomed plates as necessary, and small molecules (Section 2.14) were added if required. Plates were left at 37°C for later use. Furthermore, buffer R and buffer E (100µl electroporation kit, Thermo Fisher) were allowed to warm to room temperature.

Upon confluency, ependymal progenitor cells were rinsed in PBS and detached with TrypLE Express (Thermo Fisher Scientific, catalogue number: 12605010). An equal volume of growth media was added, and the cell suspension was centrifuged at 150 x g for 5 minutes. The supernatant was removed and the cell pellet was resuspended in PBS. A cell count was taken using a Scepter 2.0 Cell Counter (Merck). $1.5x10^5$ cells were electroporated per well of a 24-well plate. Based on this, an appropriate volume of cell suspension was transferred to a fresh Falcon tube and centrifuged at 150 x g for 5 minutes. During centrifugation, plasmids and repair templates were aliquoted into microcentrifuge tubes; 500ng of plasmid per transfection, and if included, 0.75µl of ssODN at 10µM per transfection.

Following centrifugation, cell pellets were resuspended in an appropriate volume of buffer R to bring the total volume to 10µl per reaction, and cells were mixed with the DNA. DNA-cell solutions were taken up into 10µl Neon tips, and electroporated with the following parameters: 1300V, 20ms, 2 pulses. Electroporated cells were immediately returned to growth media. Plates were then returned to the incubator. Upon confluency cells were rinsed in PBS and serum-starved to induce ciliation (Section 2.11.2).

2.12.1.3 Protocol for Neon RNP Transfection (MEFs)

The protocol for RNP Transfection of MEFs was adapted from the section 'Transfection of HEK293 FT cells using the Neon Transfection System' in the User Guide for the GeneArt Platinum Cas9 Nuclease (catalogue number B25641) from Thermo Fisher Scientific (2015). The only alterations are the potential inclusion of a repair template in step 2, and that electroporation was performed using the parameters optimised for MEFs (1350V, 30ms width, 1 pulse) in step 11.

2.12.2 4D-Nucleofector System

Transfections with the 4D-Nucleofector System (Lonza) were performed using the P4 Primary Cell 4D-Nucleofector® X Kit L (catalogue number: V4XP-4024) and the following protocol, adapted from the manufacturer's supplied protocol for MEFs (available at http://bio.lonza.com/go/op/335).

The supplied nucleofector solution was allowed to warm to room temperature. MEF growth media (Section 2.22.3) was added to the required number of wells in a six well plate, which was returned to a 37°C incubator. MEFs were dissociated with Trypsin/Versene (Section 2.22.3), and an equal volume of MEF growth media was added. The cell suspension was centrifuged at 150 x g for 5 minutes, the supernatant removed, and the pellet resuspended in 10ml of PBS. The cell suspension was diluted 1 in 2 in PBS and a cell count was performed using a Scepter 2.0 Cell Counter (Merck). Based on the cell count, an appropriate volume of the cell suspension (containing $0.5x10^5$ cells per well required) was transferred to a Falcon tube. This was centrifuged at 150 x g for 5 minutes.

Following centrifugation, the supernatant was removed and the pellet was resuspended in 100µl Buffer P4 (Lonza). The cell suspension was then transferred to the nucleocuvette vessel(s). The desired substrates (e.g. plasmid DNA) were added to the nucleocuvettes with gentle mixing. 0.5µg of plasmid was used per $1x10^5$ cells along with 0.5µl of 10µM ssODN repair template. Nucleocuvettes were gently tapped to settle the contents and nucleofected using programme CZ-167 in the X-Unit of the 4D-Nucleofector System. Following nucleofection, nucelocuvettes were incubated at room temperature for 10 minutes. Subsequently, 500µl of growth media was added to the nucleocuvette and cells were gently resuspended. An appropriately sized aliquot

of cells was transferred to the required number of wells of a 6-well plate, and plates were returned to a humidified incubator at 37°C, 5% CO_2, 3% O_2.

2.12.3 jetPRIME Transfection

jetPRIME (Polyplus transfection, catalogue number: 114-07) transfections were performed as per the manufacturer's instructions (https://www.polyplus-transfection.com/wp-content/uploads/2015/08/Short-Protocol-jetPRIME-DNA-vF.pdf). Briefly, for each well of a 6-well plate, DNA was mixed with jetPRIME reagent at a 1:2 ratio and 200µl of jetPRIME buffer was added. This mixture was vortexed and centrifuged briefly, then incubated at room temperature for 10 minutes. Subsequently, the transfection mix was added to the cells in 2ml of MEF growth media. Twenty-four hours later, media was replaced.

2.12.4 Viral Transduction

The reagents used for viral transduction of *mTmG* MEFs, along with the following protocol, were both received from Dr. Peter Tennant. Lentivirus was generated by Pam Brown (Shared University Research Facilities, University of Edinburgh), while adeno-associated virus (AAV) was generated by Christina McClure (Centre for Integrative Physiology, University of Edinburgh).

On day one, MEFs were seeded into a 24-well plate at a density of $1x10^5$ cells per well. The next day wells were generally 80-90% confluent. A fresh stock of polybrene was made in PBS to a concentration of 5mg/ml. Polybrene mix was added to MEF growth media (Section 2.22.3), 2µl per ml, and 1ml polybrene-media mix was then transferred per reaction to a 1.5ml microcentrifuge tube. To this, 4µl of lentivirus and/or 4µl of AAV repair template was added per tube and 500µl was transferred to each well of the 24-well plate. Plates were then returned to a humidified incubator at 37°C, 5% CO_2, 3% O_2 for 4 hours. Subsequently, each well was washed with PBS and 500µl of fresh growth media was added. Three days later, this transduction protocol was repeated. Twenty-four hours later, the wells were washed with PBS and fresh growth media was added.

2.13 Repair Template Design

Repair templates were all designed manually, either by myself, Dr. Pleasantine Mill or Dr. Peter Tennant. ssODN 'ultramers' were ordered from Integrated DNA Technologies (IDT), plasmids were ordered from Thermo Fisher, and minicircles were created in-house by Dr. Peter Tennant, containing the repair plasmid sequence minus the bacterial backbone.

2.13.1 *mTmG* Repair Templates

For the *mTmG* system, three different ssODNs (NLS EGFP 1-3) were designed by Dr. Pleasantine Mill, alongside three different plasmids (Figure 3.4). The minicircle used in later experiments contains the repair sequence in H2Bv2.

2.13.2 *Zmynd10* Repair Templates

For the *Zmynd10* repair experiments only ssODNs were used, which all contained the wild-type reference sequence from Ensembl (ENSMUST00000010188). ssODNs were designed specifically for each of the three guides used (Figure 5.3).

2.14 Small Molecule Enhancers of HDR

Small molecules were added to media immediately prior to transfection. All molecules were received in powdered form and resuspended in DMSO to a stock concentration recommended by the manufacturer. The stock solutions were separated into several aliquots and stored at -80°C. The 6 small molecules used were:

Brefeldin A (Cambridge BioScience, catalogue number: B012-5mg),

L-755,507 (Cambridge BioScience, catalogue number: 18629-5 mg-CAY),

Mirin (Cambridge BioScience, catalogue number: 13208-5 mg-CAY),

NU-7441 (Cambridge BioScience, catalogue number: 14881-5 mg-CAY),

SCR7 (XcessBio, catalogue number: M60082-2s) and

RS-1 (Cambridge BioScience, catalogue number: B1118-5)

Note: In 2016, Tocris indicated (https://resources.tocris.com/pdfs/archive/5342.pdf) that the correct identity of the molecule being used as SCR7, and which was received

from XcessBio, is SCR7 pyrazine. Nevertheless, it is SCR7 pyrazine which was shown to increase rates of HDR. In this thesis, 'SCR7' is taken to refer to SCR7 pyrazine exclusively.

2.15 Confocal Imaging

Imaging experiments were performed using different platforms depending on the aims of the experiment. Thirty minutes prior to most experiments, one drop of NucBlue Live ReadyProbes Reagent (Thermo Fisher, catalogue number: R37605) was added to stain nuclei. All confocal imaging experiments were carried out in glass-bottomed plates.

Platforms used include Nikon A1R and A1 confocal microscopes, Dragonfly spinning disk confocal (Andor) and Celldiscoverer 7 (Zeiss). The most up-to-date software recommended by the manufacturer was used to operate the microscopes and acquire images during each experiment. Lasers/LEDs were set up to visualise the tdTomato (561nm excitation), EGFP (488nm) and Hoechst (405nm) channels and used at the lowest possible intensities which produced a clear signal. Laser/LED intensities and gain were kept constant between all wells during an individual experiment. Depending on the aims of the experiment, a range of different magnifications was used. Generally, no environmental control, apart from heating the chamber to 37°C, was required for short imaging experiments (< 3 hours).

In order to remove bias from imaging experiments which were to be quantified, several positions (usually 10) were chosen at random per well. These were programmed, including perfect-focus settings, prior to the experiment and saved on the relevant software. This allowed automated, un-biased capture on the experimental plate.

2.15.1 Time-lapse Imaging

Time-lapses were captured on the Dragonfly spinning disk confocal (Andor). One 10µm Z-stack was captured at each position every 20 minutes. Positions were programmed prior to imaging and saved. Movement between positions was automated and the Nikon Perfect Focus System (PFS) was used to maintain focus. During time-lapses, cells were maintained in a humidified incubation chamber at 37°C, 5% CO_2.

2.16 Image Processing & Analysis

Images were processed using the appropriate software for each imaging platform, that being Fiji/ImageJ (Schindelin, Arganda-Carreras et al. 2012, Rueden, Schindelin et al. 2017) for Nikon platforms, Imaris (Bitplane) for Dragonfly and Zen (Zeiss) for the Celldiscoverer 7. Where post-processing was applied, such as changes in brightness or contrast, it was applied equally to all images in the experiment.

2.16.1 Cell Counter and Nucleus Counter (Fiji)

To calculate the number of cells in an image, the Nucleus Counter plugin in Fiji was employed. Settings were optimised for the Hoechst stain levels in each experiment and were validated by manually counting three images and ensuring Nucleus Counter results were within 15% of the manual counts. GFP-positive nuclei were manually counted using the Cell Counter plugin in Fiji.

2.16.2 Imaris Time-lapse Analysis Settings

Imaris (Bitplane) was used with the 'Track' module enabled to measure changes in fluorescence intensity over time. The Track module settings used to measure and track EGFP, tdTomato and background intensities are included on the attached media CD (ImarisTrackAnalysisSettings.rtf). 'Tomato Removal' settings were applied prior to measuring the background intensity, to 'remove' cells from an image, leaving background regions.

2.16.3 CellProfiler & CellProfiler Analyst

CellProfiler (Carpenter, Jones et al. 2006) and CellProfiler Analyst (Jones, Kang et al. 2008) were used to count nuclei and GFP-positive nuclei in images. CellProfiler was programmed to identify nuclei, in the Hoechst channel. Following this, the 'machine-learning' tool, CellProfiler Analyst, was 'taught' on an image-by-image basis to identify GFP-positive nuclei by manually selecting several positive cells. CellProfiler Analyst then analysed the entire image, providing a total count.

2.17 Fluorescence-Activated Cell Sorting

Cells were prepared for flow cytometry (FC) or fluorescence-activated cell sorting (FACS) by rinsing in PBS and dissociating thoroughly in Trypsin/Versene (Section 2.22.3) at 37°C for 5-10 minutes. Cells were aspirated by pipette several times to ensure a single cell suspension had been created. Following this, an equal volume of growth media was added and cells were transferred to 15ml Falcon tubes. These were centrifuged at 150 x g for 5 minutes, and the cell pellet was resuspended in PBS, generally 200-500µl. This cell suspension was transferred into a 4.5ml FACS tube (BD Falcon).

To provide cell counts, a BD LSRFortessa or BD Accuri C6 was used, whereas to sort cells a BD FACSJazz or BD FACSAria were used (All BD Biosciences). Sorting gates were manually set on all machines. For EGFP an excitation filter of 488/50 was used with an emission filter of 525/50 (488-525/50). For Tomato, 561-610/20, 561-586/15 or 561-582/15 were used depending on the machine. Within an experiment gates were kept constant between samples to allow comparison. For population counts, generally 100,000 cells were sorted per sample.

2.18 Next-generation Sequencing

Next-generation sequencing (NGS) was carried out on two platforms, the Ion Torrent Personal Genome Machine (PGM) System (Thermo Fisher) using the Ion Torrent 314 v2 chip (400bp reads), or the Oxford Nanopore MinION.

2.18.1 Ion Torrent

Samples were prepared for Ion Torrent sequencing either by myself or by the Edinburgh Wellcome Trust Clinical Research Facility (WTCRF). Sample preparation for MinION runs was performed by the WTCRF.

2.18.1.1 Ion Torrent In-house Sample Processing

For the *Zmynd10* Ion Torrent sequencing run (Section 5.5), PCR samples were prepared for sequencing by myself and then sequenced at the IGMM. PCR products were purified twice using Agencourt AMPure XP beads (Beckman Coulter) as per the manufacturer's instructions. Purified PCR products were analysed on an Agilent

Bioanalyzer 2100 using a high sensitivity DNA assay. Following this, an equimolar pool of all samples was prepared and submitted to the IGMM technical services for sequencing.

2.18.1.2 Ion Torrent Processing by the WTCRF

PCR Products were purified (Section 2.4.2) and eluted in TE-buffer before being transferred to the WTCRF. Samples were then processed by the facility. Amplicons were quantified using a Qubit dsDNA HS kit (Thermo Fisher, catalogue number Q32851) and then sheared using a Covaris E220 Evolution Focused Ultrasonicator (Thermo Fisher). Sheared samples were quantified, and Ion-compatible barcoded adapters (Thermo Fisher) were attached. The library was then amplified (10 cycles) and size selected using AMPure XP beads (Beckman Coulter) for fragments approximately 300bp in length. Barcoded libraries were checked for purity and quantified, and an equimolar stock was prepared and sequenced.

2.18.2 MinION Processing by the WTCRF

Amplicons were quantified using a Qubit dsDNA BR assay (Thermo Fisher, catalogue number: Q32850) and 50ng of each amplicon was end-repaired and adenylated using an NEBNext Ultra End Repair/dA-Tailing Module kit (NEB, catalogue number: E7442S) and purified using AMPpure XP beads (Beckman Coulter). Barcode adapters from the PCR Barcoding Kit 96 (Oxford Nanopore Technologies, catalogue number: EXP-PBC096) were ligated to the end-repaired, dA-tailed DNA during 18 cycles of PCR. Excess barcode adapters were removed using AMPure XP beads, and barcoded DNA was quantified using a Qubit dsDNA HS assay (Thermo Fisher, catalogue number Q32851). Equal quantities of each barcoded amplicon were pooled before being end-repaired and adenylated to allow ligation of sequencing adapters and tethers from the Nanopore $1D^2$ Sequencing Kit (Oxford Nanopore Technologies, catalogue number: SQK-LSK308). Libraries were purified with AMPure XP beads before elution in a proprietary Running buffer. The sequencing library was quantified using a Qubit dsDNA HS assay and an equimolar stock was prepared and sequenced.

2.19 Sequencing Analysis Pipelines

Pipelines were designed in-house to analyse the resulting data from NGS and were created as shell scripts in Unix which call on various software packages to perform

tasks such as alignments and quality control. The pipeline was initially designed for Ion Torrent data, and then adapted to process MinION data. The scripts used are provided on the attached media CD (IonTorrent_Editing_Analysis.sh and MinION_Editing_Analysis.sh).

2.19.1 Ion Torrent Analysis Pipeline

The Ion Torrent script takes as input a fastq file, as is output from an Ion Torrent run, and uses Bowtie 2 (Langmead and Salzberg 2012) to align reads against a custom reference sequence which is also supplied by the user. BamQC (Simon Andrews, https://github.com/s-andrews/BamQC) is called to provide quality control metrics on the input data. Following this, samtools (Li, Handsaker et al. 2009), bam-readcount (https://github.com/genome/bam-readcount) and the Genome Analysis Toolkit (McKenna, Hanna et al. 2010) packages are used to generate alignment statistics. Two different variant callers are invoked for comparison, VarScan 2 (Koboldt, Zhang et al. 2012) and the Bcftools package by samtools. Bowtie 2 alignments were visualised using the Integrative Genomics Viewer (Robinson, Thorvaldsdóttir et al. 2011, Thorvaldsdottir, Robinson et al. 2013), which was also used to produce several figures.

2.19.2 MinION Analysis Pipeline

The MinION processing script was derived from the Ion Torrent script and is largely the same except that GraphMap (Sovic, Sikic et al. 2016) is used to align reads and that the following alignments are 'cleaned up' using Picard Tools (http://broadinstitute.github.io/picard). GraphMap contains a dedicated algorithm for aligning Oxford Nanopore data. Prior to running the MinION script, the .fast5 output was converted to .fastq using Poretools (Loman and Quinlan 2014), and then processed using Porechop (https://github.com/rrwick/Porechop) to split the file by barcode. A BBMap script, readlength.sh (https://github.com/BioInfoTools/BBMap), was used to generate read length histograms and calculate mean/median read lengths.

2.20 *De novo* Genome Assembly

SPAdes (Bankevich, Nurk et al. 2012) was used to assemble Ion Torrent data, whereas Canu (Koren, Walenz et al. 2017) was used to assemble MinION data. Settings in both

were tailored to expect a small, repetitive genome. The programme Bandage (Wick, Schultz et al. 2015) was used to visualise the resulting genome assemblies.

2.21 Graphing & Statistical Analyses

Counts from imaging and FACS/FC were plotted graphically and analysed using the Microsoft Excel package. Where biological replicates were performed, a Student's t-test was performed to determine the statistical significance of any difference between populations. T-test results below p=0.05 were considered to be statistically significant. Furthermore, for experiments with biological replicates, graphs were plotted as the mean ± 1 SD to show variation in the results, unless otherwise stated. In cell culture experiments with the *mTmG* line, each batch of cells (biological replicate) was treated (generally transfected with or without drug), maintained and assayed separately for the entire experiment.

2.22 List of Recipes

2.22.1 General Recipes

20x TBE (1000ml):

216.0 g Tris base

110.0 g Boric Acid

80mL 0.5 M EDTA, pH 8.5

Add ddH_2O to 1000ml

1x PBS (1000ml):

800ml ddH_2O

8.00g NaCl

0.20g KCl

1.44g Na_2HPO_4

0.24g KH_2PO_4

Adjust pH to 7.4 with HCl

Add ddH_2O to 1000ml

10x Orange G Loading Dye (50ml):

30% Glycerol in ddH_2O

100mg Orange G

1x Surveyor Buffer:

10mM Tris-HCl, pH 8.8

15mM $MgCl_2$

50mM KCl

2.22.2 Recipes for Bacterial Culture

Luria-Bertani Broth (LB) (1000ml):

10.00g Tryptone

5.00g Yeast Extract

10.00g NaCl

1.00g Glucose

Add ddH_2O to 1000ml

Add Ampicillin (to 100µg/ml) or Kanamycin (to 50µg/ml) as required

Freezing Media

70% Luria Broth

30% Glycerol

LB-Agar (1000ml):

10.00g Tryptone

5.00g Yeast Extract

10.00g NaCl

1.00g Glucose

15.00g Agar

Add ddH_2O to 1000ml

Add Ampicillin (to 100µg/ml) or Kanamycin (to 50µg/ml) as required

2.22.3 Recipes for Tissue Culture

MEF Growth Media (500ml):

450ml Opti-MEM (Thermo Fisher, catalogue number: 31985070)

50ml FCS

5ml P/S

3.7µl β-mercaptoethanol

MEF Freezing Media (50ml):

35ml MEF growth media

10ml FCS

5ml DMSO

Hank's Solution (100ml):

84ml Sterile, filtered tissue culture water

10ml HBSS + Ca + Mg (Thermo Fisher, catalogue number: 24020083)

5ml 1M HEPES (Thermo Fisher, catalogue number: 15630056)

1ml P/S

Ependymal Growth Media (500ml):

450ml DMEM + GlutaMAX (Thermo Fisher, catalogue number: 10566016)

50ml FCS

5ml P/S

Ependymal Serum-Starvation Media (500ml):

495ml DMEM + GlutaMAX (Thermo Fisher, catalogue number: 10566016)

5ml FCS

5ml P/S

1x Trypsin (1000ml):

2.00g Trypsin 1:250

0.06g Penicillin

0.13g Streptomycin

5.0ml 0.2% Phenol Red

Add 1x PBS to 1000ml

Adjust pH to 7.8 with $NaHCO_3$

1x Versene (1000ml):

0.40g Sodium EDTA

5.00ml 0.2% Phenol Red

Add 1x PBS to 1000ml

Trypsin/Versene (100ml):

50ml 1x Trypsin

50ml 1x Versene

Quick Lysis Buffer (500ml):

25ml 2M Tris-HCl pH7.5

5ml 0.5M EDTA

10ml 10% SDS

20ml 5M NaCl

440ml ddH_2O

Chapter 3 Repurposing of the *mTmG* Fluorescent Reporter System for CRISPR-mediated Genome Editing

3.1 Introduction

3.1.1 Repurposing of the Lineage-Tracing *mTmG* Fluorescent Transgenic Mouse

Fluorescent reporters allow a rapid visual read-out of cellular events, both at a cellular and organismal level, which can be visualised and quantified using imaging or via flow cytometry (FC). Several fluorescent reporter systems have been made to track genome recombination events by Cre-recombinase, including the Z/EG reporter mouse (Novak, Guo et al. 2000). This is a random insertion transgenic mouse line, where a floxed *LacZ* is driven by the *Cytomegalovirus* (CMV) enhancer/chicken actin promoter (pCA); which while widely expressed is not ubiquitous, including in the airways. When crossed onto a Cre-deleter strain, *lacZ* expression is replaced with *GFP*. In another line from the Constantini group, the pCA promoter was used to drive a floxed CFP or YFP, this time targeted to the ubiquitously expressed *Rosa26* locus, marking each cell type in the embryo and adult (Srinivas, Watanabe et al. 2001). The *mTmG* mouse designed by Muzumdar, Tasic et al. (2007) contains a double fluorescence Cre-reporter cassette, which is also targeted to the *Rosa26* locus. In the absence of Cre-mediated recombination, the ubiquitously expressed pCA promoter drives expression of a floxed membrane-tdTomato-polyA sequence. Upon Cre-mediated recombination, the Tomato portion is excised, and the pCA promoter drives ubiquitous expression of membrane-EGFP (Figure 3.1).

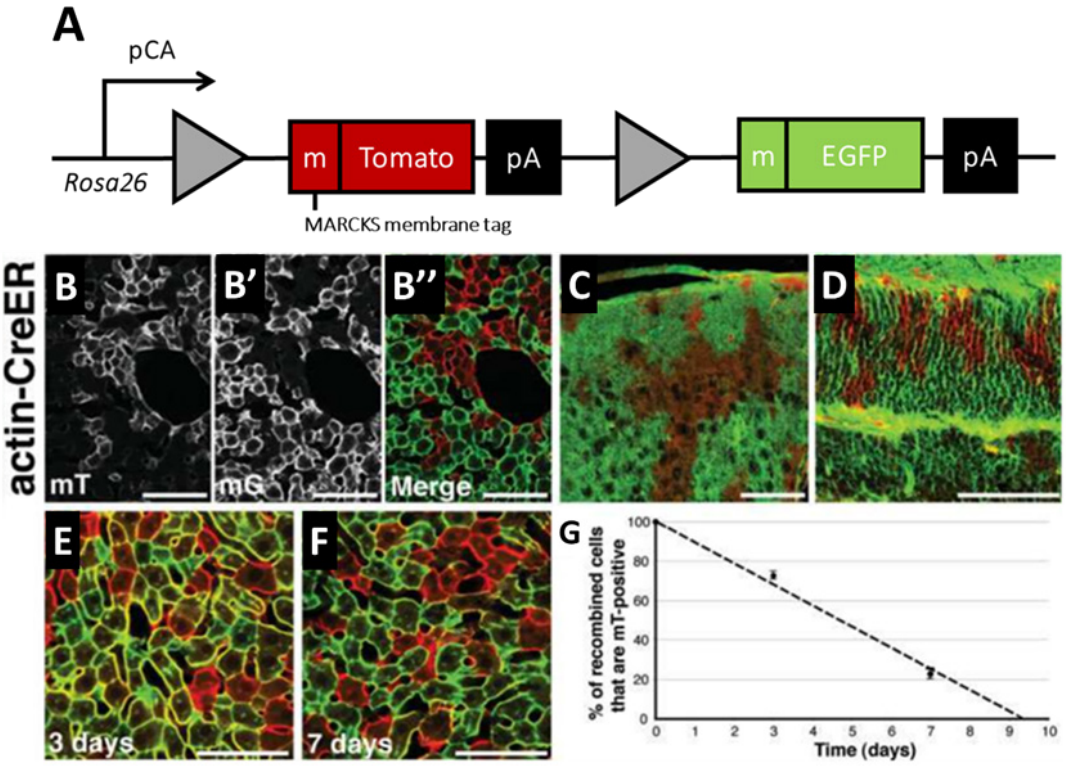


Figure 3.1: Cre-mediated recombination in the *mTmG* mouse leads to a specific transition from membrane-Tomato to membrane-GFP. A) Schematic of the *mTmG* locus. Grey triangles represent loxP sites. pA = polyadenylation signal. Not to scale. B–D) Following injection with tamoxifen at P7, *actin-CreER* mice display a complementary pattern of mTomato and mEGFP labeling in the liver (B-B''), cerebral cortex (C), and retina (D). E-G) Following injection with tamoxifen, liver hepatocytes from adult *mTmG;actin-CreER* mice can be seen to grandually transition from mTomato to mEGFP. The persistence of membrane-Tomato is evidenced by the double-positive cells in (E) which can be seen to have largely disappeared in (F). A linear regression indicates a total elimination of mTomato after ~9 days (G). Scale bar: (B–F) 50 µm. Figure adapted from Muzumdar, Tasic et al. (2007).

Given the nature of this dual reporter, we believed the *mTmG* system could be adapted to provide a visual read-out of genomic editing events. By targeting CRISPR/Cas9 to an identical sequence which surrounds the two loxP sites, we can distinguish non-edited cells (membrane Tomato) from edited cells (other); allowing us to quantify editing efficiencies. Furthermore, by changing the localisation signal from a MARCKS membrane tag to a nuclear signal, we can separate both non-homologous end joining (NHEJ, membrane EGFP) and homology-directed repair (HDR, nuclear EGFP) driven by CRISPR/Cas9 by supplying a repair template. The *Rosa26 mTmG*

locus, along with our system for measuring rates of editing and repair, is illustrated in Figure 3.2:

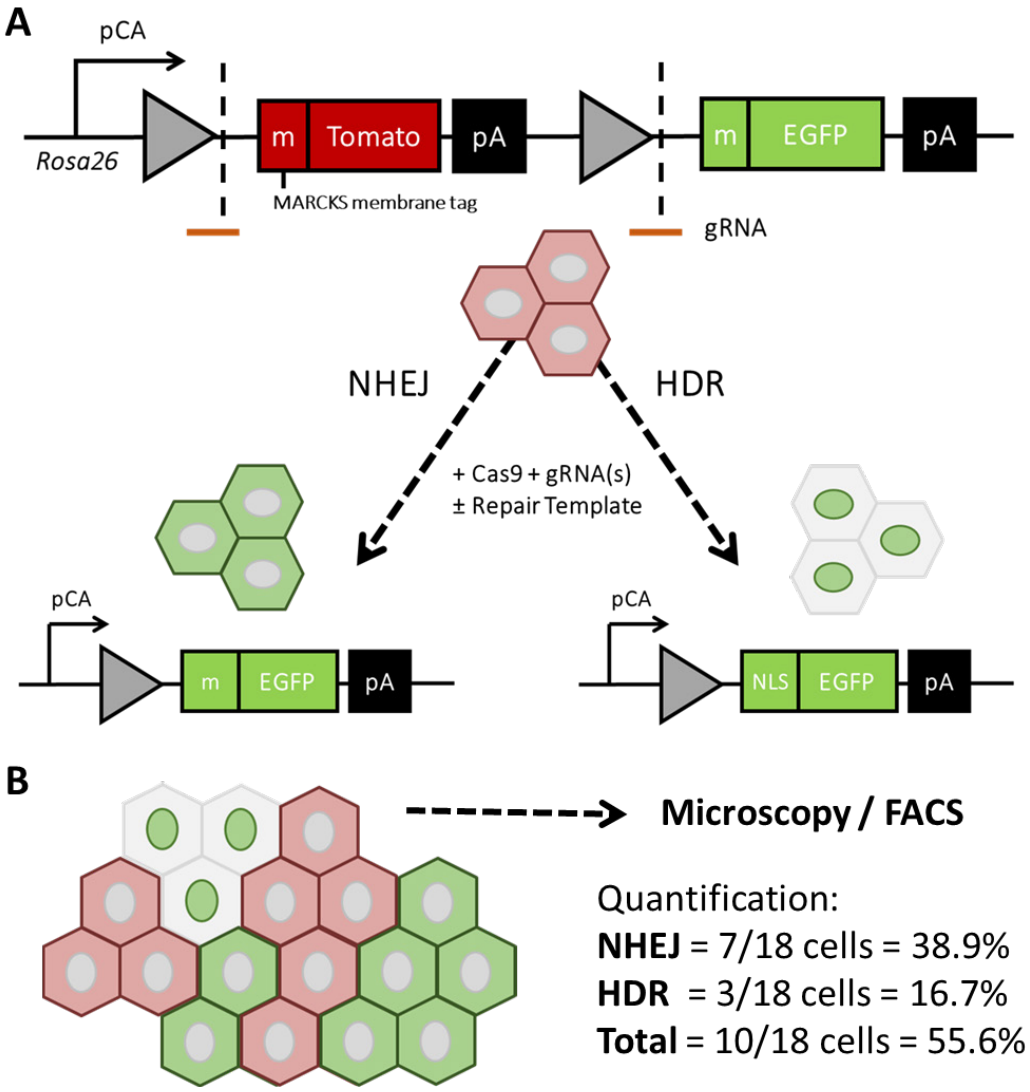


Figure 3.2: Schematic representation of the repurposed *mTmG* fluorescent reporter system, for measuring CRISPR-mediated genome editing (NHEJ) and repair (HDR) A) On the *Rosa26* mTmG locus the target sequence for the guide-RNA (gRNA) is indicated with an orange bar, dashed lines Cas9 cleavage sites, the black arrow indicates the promoter, grey triangles loxP sites, boxes genomic elements and hexagons cells with oval nuclei inside. B) The resulting mixed population following editing can be quickly quantified using imaging or FC/FACS to give rates of NHEJ and HDR.

3.2 Design Strategy for Repurposing the *mTmG* Locus to Report on Genome Editing Events

3.2.1 Design of Guide-RNAs

Two guides were designed in the laboratory to target the identical sequence surrounding the loxP sites in the *Rosa26 mTmG* locus (Section 2.1). Both guides were cloned into the Cas9 expression vector pX330 (Section 2.2). Using either of these guides, we would expect the creation of two double strand breaks (DSBs) resulting in the loss of the intervening 2.4kb fragment containing the membrane-tagged Tomato (Figure 3.3). The use of two different guides allows us to compare the efficiency of deletion at the same locus, with different target sequences.

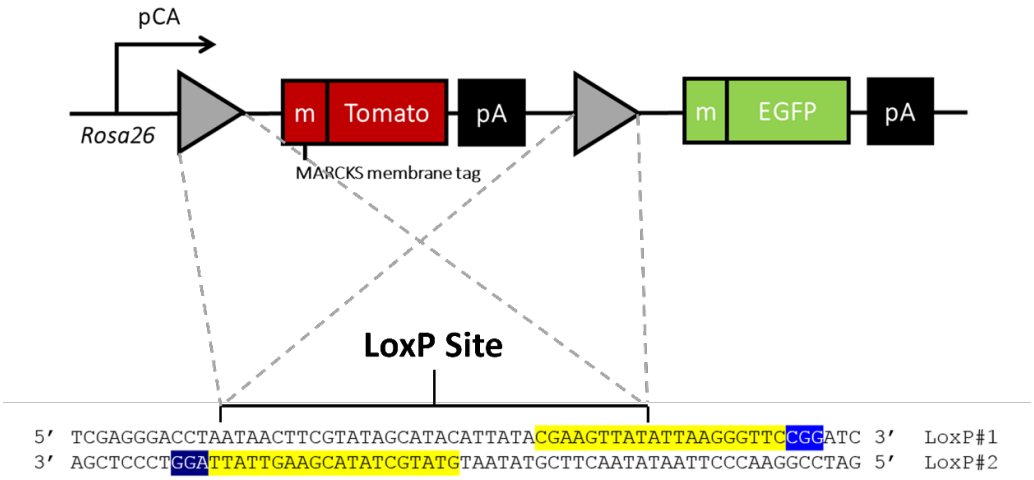


Figure 3.3: Schematic showing the location of the two guides used to target the loxP sites in the *mTmG* locus. Above: black arrow represents the promoter, grey triangles loxP sites and boxes genomic elements. Below: yellow indicates the target site while blue indicates the PAM.

3.2.2 Repair Template Design

As the ideal format for repair is not currently known, six different repair templates were designed in the lab (Section 2.13) and employed in the following work, to attempt to maximise rates of HDR. These repair templates have varying sized regions of homology, and were in various formats, such as single-stranded oligonucleotide (ssODN), plasmid and minicircle (a plasmid with the bacterial backbone removed). These repair templates were all designed to replace the MARCKS membrane tag with a nuclear signal; either an SV40 NLS sequence or the (well tolerated) human histone

H2B sequence, to determine which provides cleaner nuclear localisation. Whereas the NLS sequence requires active transport into the nucleus, proteins tagged with H2B are robustly incorporated into nucleosomes in cells from mice at all developmental stages (Kanda, Sullivan et al. 1998, Hadjantonakis and Papaioannou 2004). Furthermore, all but one of the repair templates harboured additional changes in that the PAM sites, necessary for Cas9 binding, in order to prevent cleavage of the template and repaired alleles. The repair templates used are summarised in Figure 3.4:

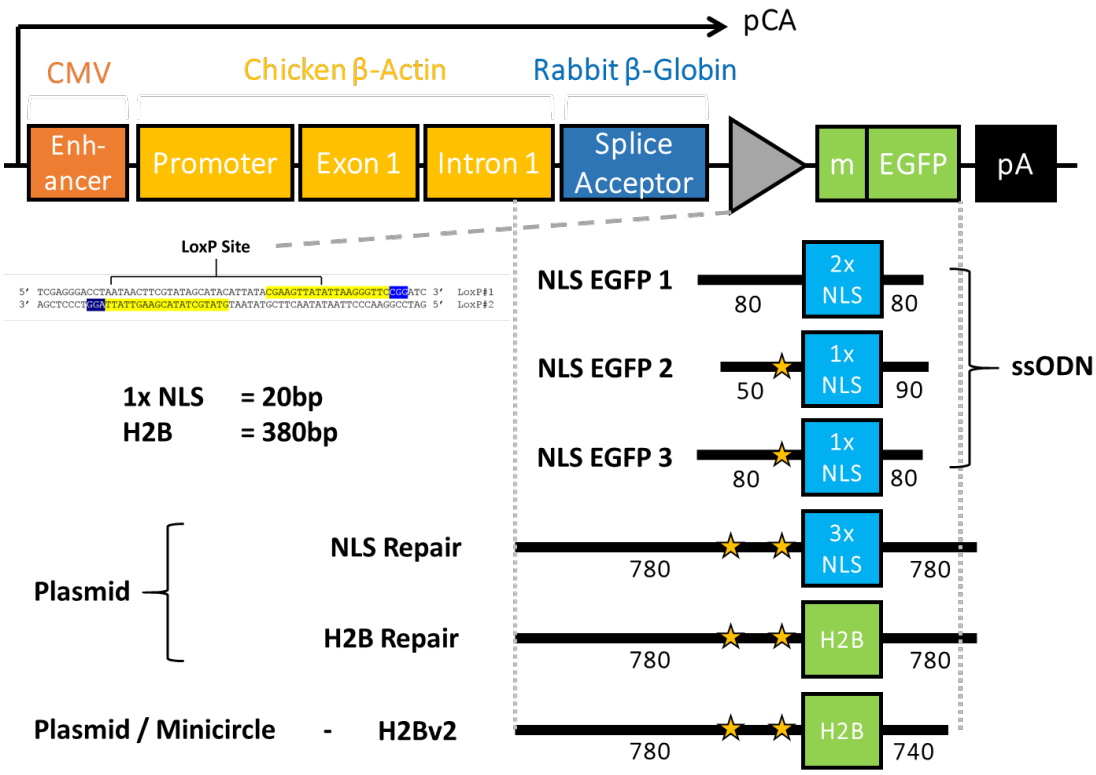


Figure 3.4: Schematic representation of the repair templates used, displaying their regions of homology with respect to the pCA promoter and EGFP sequence. The size of the homology arms (in nt/bp) are indicated. Single basepair changes which abolish the PAM sites for LoxP#1 and LoxP#2 guides are indicated with yellow stars. H2Bv2 lacks the final ten codons of the EGFP sequence. Above: black arrow represents the promoter, grey triangles loxP sites and boxes genomic elements. Not to scale.

3.3 Derived *mTmG* MEFs and Ependymal Cells Display Stable Membrane-Tomato Fluorescence

To test that the *mTmG* reporter system could be utilised to monitor genome editing events, we isolated primary cells from the *mTmG* mouse. Mouse embryonic fibroblasts (MEFs) and ependymal cells were derived from E11.5 and E18.5 mice, as described in Sections 2.10.1 and 2.11.1 respectively. Cells were cultured until a relatively homogenous population was obtained, 7 days for ependymal cells or passage 3 (P3) for MEFs. The ependymal radial glial progenitors were then serum-starved to induce terminal differentiation and multi-ciliation, forming motile ciliated ependymal cells. Both cell types were imaged following this process on a Nikon A1R confocal microscope at 20x, which confirmed that all cells display membrane-Tomato fluorescence with no cells displaying membrane-EGFP or loss of fluorescence. This indicates that there is no spontaneous conversion of the locus in the absence of Cas9 or Cre (Figure 3.5).

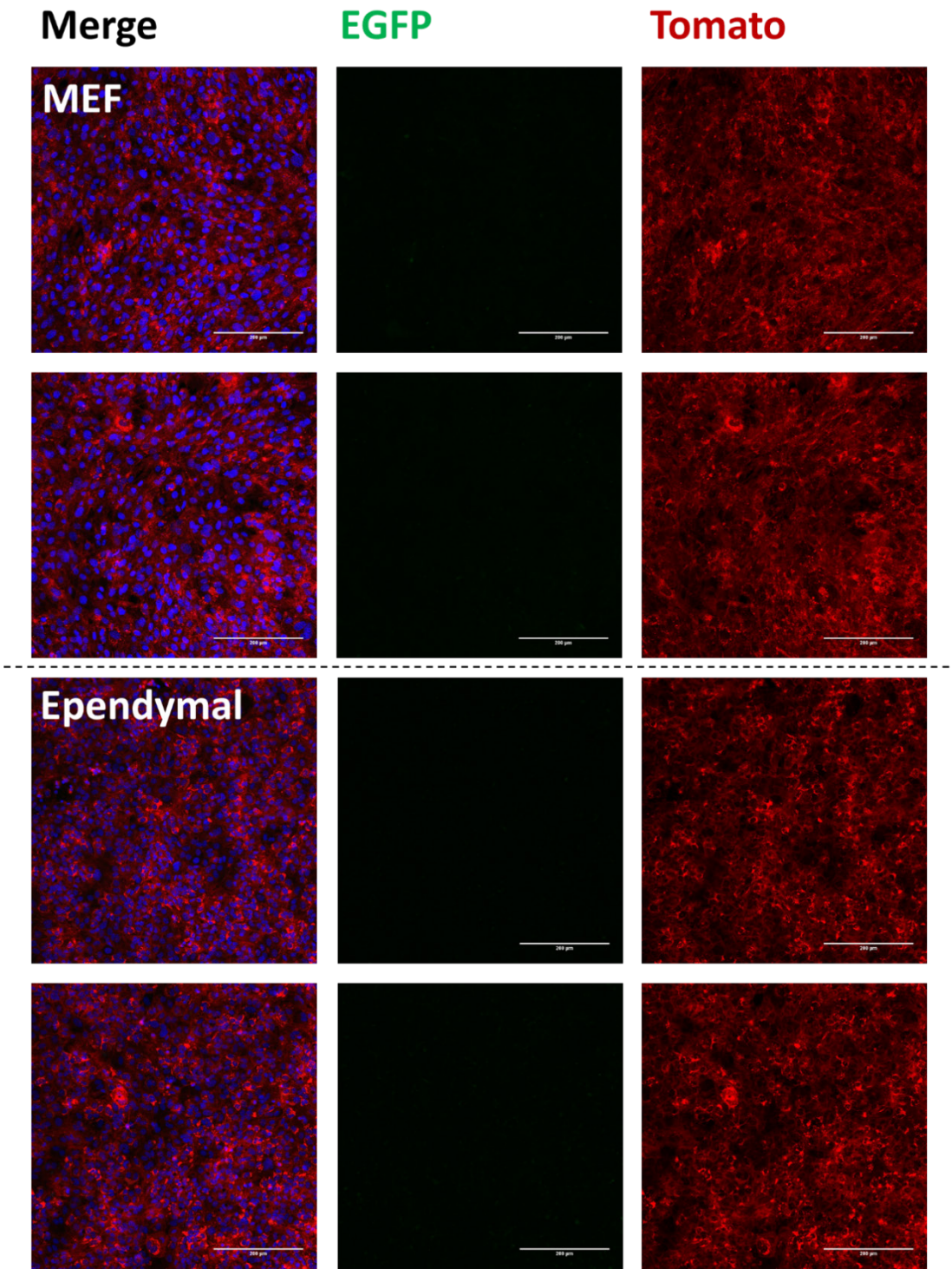


Figure 3.5: Unedited *mTmG* MEFs (above) and *mTmG* ependymal cells (below) display only Tomato expression. Cells were imaged at 20x on a Nikon A1R confocal microscope. Scale bar 200µm.

3.4 Optimising Delivery of the Editing Machinery to Primary Cells

In order to test that our system reports on genome editing, I transiently transfected our two cells lines; MEFs, with primary cilia only, and ependymal cells, a more clinically relevant motile-ciliated cell type, using a range of nucleofection platforms; including the Neon Transfection System (Thermo Fisher) and the 4D-Nucleofector System (Lonza). Furthermore, the lipid-based jetPRIME transfection reagent (Polyplus) was tested. All initial transfections were performed with the LoxP#1 pX330 plasmid alone. Further details on the transfection protocols can be found in Section 2.12. Following these trials, the Neon Transfection System, was found to be the most consistently efficient method of transfection, for both MEFs and ependymal cells. The 4D-Nucleofector system resulted in a high amount of cell death, whereas transfection rates were very low with the jetPRIME transfection reagent (not quantified, images not shown).

3.4.1 Neon Transfection System – MEFs

mTmG MEFs were transfected with LoxP#1 pX330 using the Neon Transfection System as described in Section 2.12.1 with 10µl transfection tips. As controls a non-electroporated and no DNA (electroporated) control were also prepared. 120-hours post-transfection, cells were imaged on a Nikon A1R confocal microscope. A large number of cells in the transfected population could be seen to have gained EGFP-fluorescence, and many of the same cells appeared to have lost Tomato fluorescence. These cells were then analysed by FC at 144-hours post-transfection, at which time ~20% of the cells appeared to be edited (Figure 3.6). The edited cells fell broadly into four populations; Tomato-positive (Q1), double-positive (Q2), double-negative (Q3) and GFP-positive (Q4). The double-positive (GFP^+ $Tomato^+$) population could potentially be explained by the ~4.5 day half-life of tdTomato (Muzumdar, Tasic et al. 2007) or if only one allele had been edited, resulting in both Tomato and GFP being expressed.

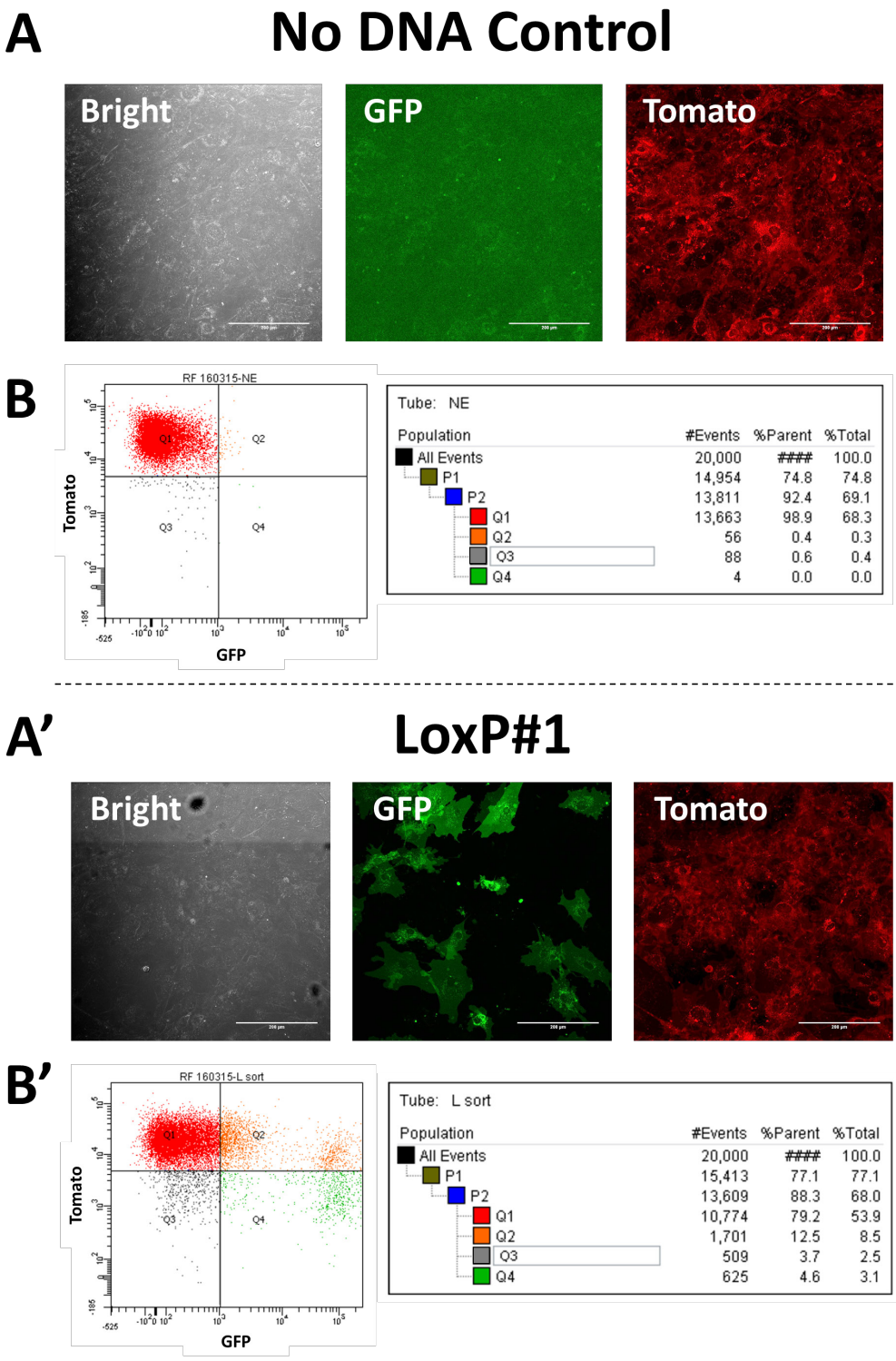
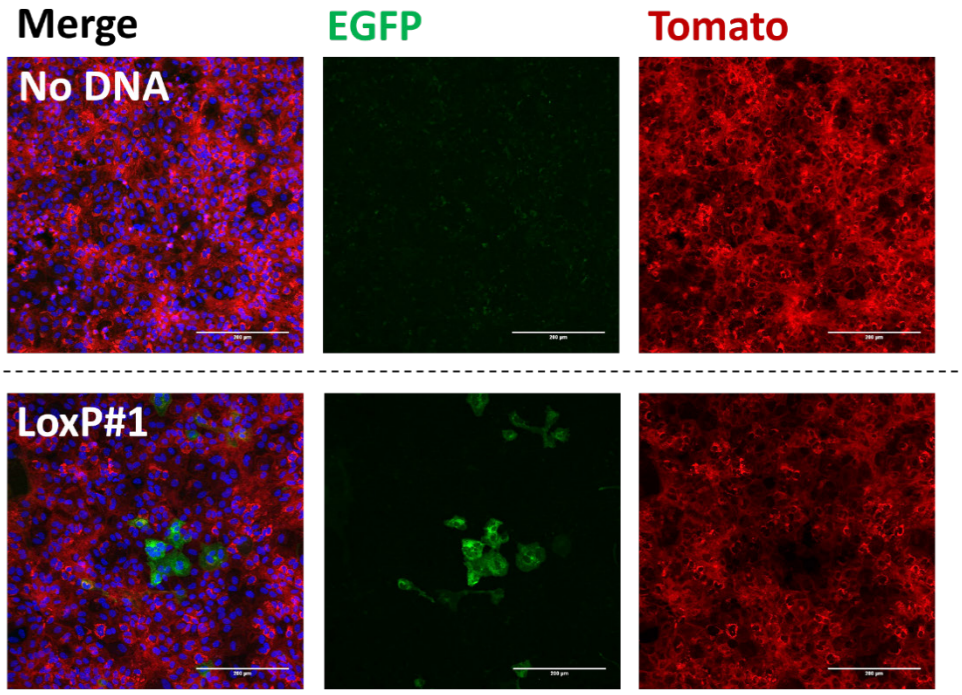


Figure 3.6: The Neon Transfection System provides high transfection efficiency (~21%) and low cell death in *mTmG* MEFs. A, A') Cells were imaged at 20x, 120-hours post transfection, at which point GFP-positive cells could clearly be seen in the transfected sample. B, B') Cells were analysed by FC 144-hours post-transfection, revealing all non-transfected cells are Tomato-positive, whereas there is a substantial shift in fluorescence in the transfected population. Scale bars 200µm.

3.4.2 Neon Transfection System – Ependymal Cells

During the expansion phase, ependymal progenitor cells were transfected with LoxP#1 pX330 using the Neon System (Section 2.12.1.2). A no DNA, electroporation control was also prepared. Transfected cells were allowed to proliferate for 5 days following transfection prior to serum-starvation. Approximately 14 days following serum-starvation, cells began to form multicilia. 18 days following serum-starvation, wells were imaged on a Nikon A1R confocal microscope. Each well was imaged in 5 random positions, and images were quantified using the Nucleus Counter and Cell Counter plugins in Fiji (Figure 3.7). GFP-positive cells, which appeared by morphology to be ependymal cells, were present in every image, although the transfection efficiency was rather low at 3.2%.

A



B

Image	Green Cells	Total Cells
1	24	745
2	23	783
3	26	805
4	37	797
5	18	853
Total	128	3983
Transfection Efficiency		3.21%

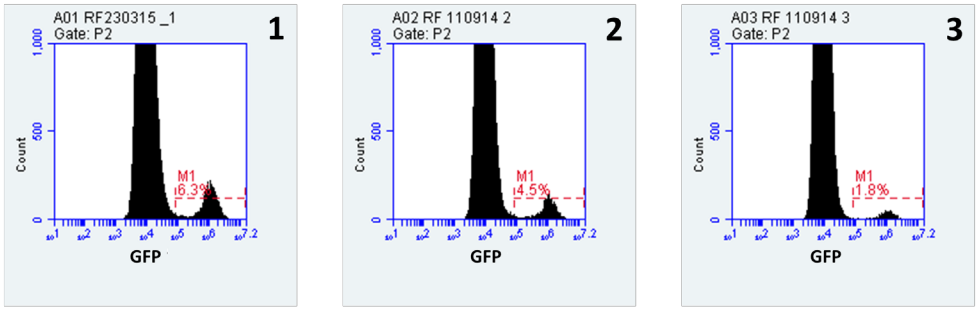
Figure 3.7: The Neon Transfection System transfects ependymal progenitor cells, albeit at low efficiency (~3.2%). A) Confocal images of non-transfected and LoxP#1 pX330 transfected ependymal cells captured at 20x, 18 days post-transfection. Scale bar 200µm. B) Nucleus Counter and Cell Counter plugins in Fiji were used to quantify the images.

3.4.3 Plasmid Concentration is a Limiting Factor in Transfection Efficiencies

Low transfection efficiencies associated with the Neon Transfection System could be due to excessive plasmid/DNA concentration and the following death of transfected cells. To test this I decided to perform a concentration curve with the LoxP#1 pX330 plasmid. *mTmG* MEFs were transfected using the Neon system and 10µl tips (Section 2.12.1). Three different concentrations of LoxP#1 pX330 were used, along with a *Zmynd10* pX330 plasmid control (targets the endogenous *Zmynd10* locus which is not expressed in MEFs), a no DNA electroporation control, and a cell only control. $3x10^5$ cells were transfected per well of a 6-well plate.

Seventy-two hours post-transfection, each well was dissociated and cells were scored for GFP fluorescence using a BD FACSAccuri C6 flow cytometer (BD Biosciences). FC results showed a linear decrease in the number of GFP-positive cells upon decreasing plasmid concentration, suggesting this is a limiting factor. Cell survival did not appear to be affected by DNA concentrations in this range, but may have been correlated to be the time spent in the transfection buffer, buffer R (Figure 3.8).

A



B

ID	DNA/reaction (ng)	Plasmid	GFP (%)	Cells/µl
1	500	LoxP#1 pX330	6.2	400
2	250	LoxP#1 pX330	4.5	291
3	125	LoxP#1 pX330	1.8	289
4	0	N/A (Electropotated)	0.0	190
5	0	N/A	0.0	110
6	500	*Zmynd10* pX330	0.0	186

Figure 3.8: There is a trend to higher transfection efficiencies with greater plasmid concentrations without cytotoxicity in *mTmG* MEFs. A) FC plots of cell count versus GFP intensity. B) Summary of transfection conditions for each well along with the percentage of GFP+ cells and cell densities at 72 hours post-transfection.

3.4.4 Scaling up Neon Transfections

In order to mitigate the effect of buffer R on cell survival, while still transfecting a large number of cells, I tested whether similar transfection efficiencies could be achieved using 100µl Neon electroporation tips instead of the 10µl tips used previously. This would allow us to transfect $5x10^5$ cells per reaction as opposed to $0.5x10^5$ (10x increase).

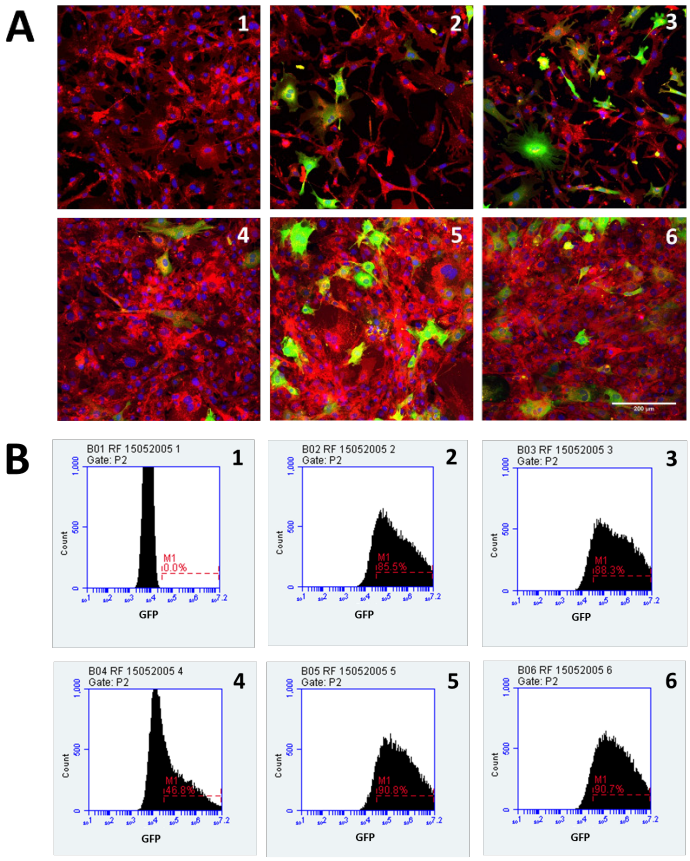
Reagents were scaled up ten-fold for the 100µl tips; 5000ng pmaxGFP plasmid (Lonza) was used per reaction, and a range of voltages was assayed. At the same time, two controls were prepared: a well transfected using the 10µl tips as previously and a

no DNA control. Furthermore, in order to see whether plasmid concentrations could be increased further, a 750ng DNA/reaction well was prepared using the 10µl tips.

Forty-eight hours post-transfection wells were imaged at 20x on the Nikon A1R (Figure 3.9A). The following day, wells were dissociated and the GFP-intensity of the cells was measured using a BD FACSAccuri C6 flow cytometer. Results indicated that the 100µl tips may even provide higher transfection efficiencies than the 10µl tips (91% vs. 86%), whereas the current voltage used appeared to be ideal (Figure 3.9C). There did not appear to be a substantial increase in transfection efficiency with the (50%) higher plasmid concentration. Based on this, it was decided to use the following transfection parameters for MEFs with the 100µl tips in all future experiments, unless otherwise stated:

1000ng of DNA per $1x10^5$ cells, 1350V, 30ms, 1 pulse

Furthermore, comparing the results of this experiment with the previous section, it is apparent that there is either a substantial difference in transfection efficiencies or cell survival following transfection with pX330 and pmaxGFP. Alternatively, these results could suggest that the majority of cells transfected with pX330 are not being edited. These two possibilities will be investigated further in the following section, and Section 3.15.



C

ID	Tip Used	DNA/reaction (ng)	Electroporation Parameters	Plasmid	GFP (%)
1	10µl	0	1350V, 30ms, 1 pulse	N/A	0.0%
2	10µl	500	1350V, 30ms, 1 pulse	pmaxGFP	85.5%
3	10µl	750	1350V, 30ms, 1 pulse	pmaxGFP	88.3%
4	100µl	5000	1300V, 30ms, 1 pulse	pmaxGFP	46.8%
5	100µl	5000	1350V, 30ms, 1 pulse	pmaxGFP	90.8%
6	100µl	5000	1400V, 30ms, 1 pulse	pmaxGFP	90.7%

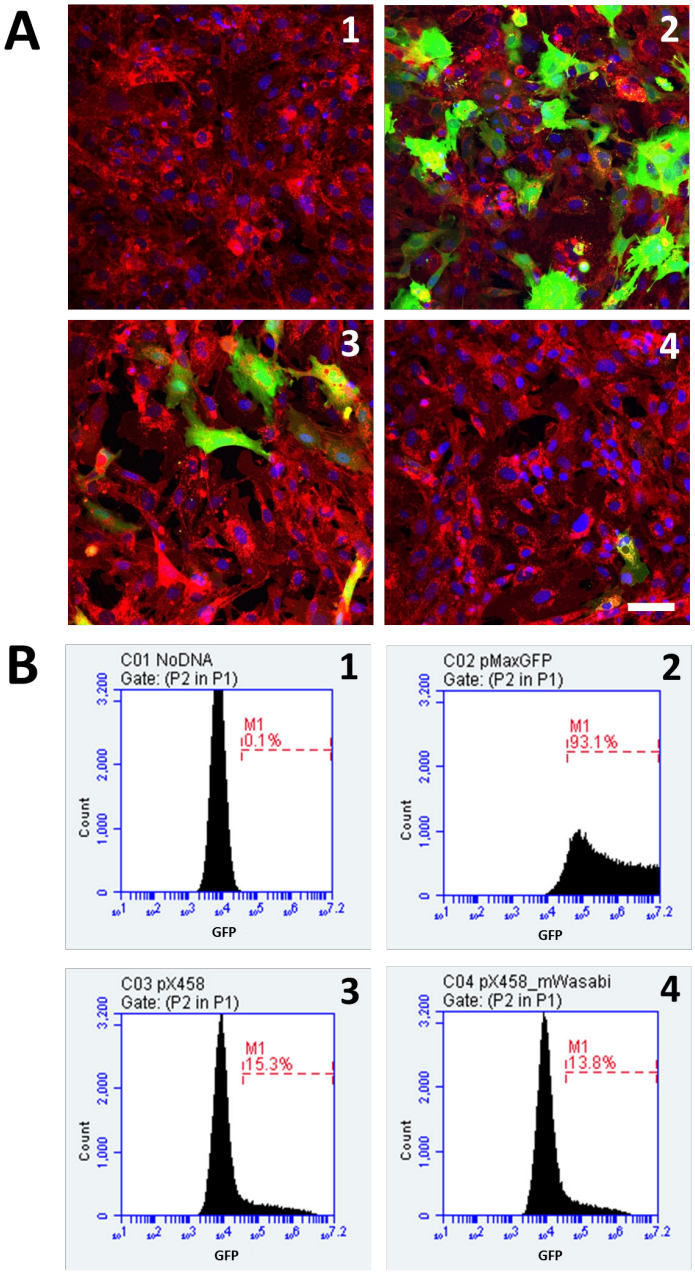
Figure 3.9: High rates of transfection (~90%) are seen with the 100µl Neon electroporation tips in MEFs A) Confocal images of each well taken at 20x, 48 hours post-transfection. Scale bar 200µm. B) FC plots of the wells showing cell count versus GFP intensity, taken at 60 hours post-transfection. C) Summary of transfection conditions used for each well along with the resulting percentage of GFP+ cells.

3.4.5 Plasmid Size may be a Limiting Factor in Transfection Efficiencies

In order to ascertain whether the size or associated cytotoxicity of the pX330 plasmid could account for the large difference in the number of GFP-positive cells between LoxP#1 pX330 and pmaxGFP, I decided to test the transfection efficiency of a larger (Cas9 containing) GFP expression vector. To do this, an empty pX458 plasmid was employed which is a similar size to pX330 (9.3kb and 8.5kb respectively).

Furthermore, a brighter variant of pX458 was included for comparison. This contained a membrane-tagged monomeric green fluorescent protein derived from mTFP1 Wasabi, cloned into the pX458 plasmid, to create pX458-mWasabi (generated in our group by Peter Budd), These vectors, including pmaxGFP, were transfected into *mTmG* MEFs, $5x10^5$ cells per well (Section 2.12.1.1). Twenty-four hours post-transfection, each well was imaged on the Nikon A1R confocal, then following imaging the cells were dissociated and analysed on a BD FACSAccuri C6 flow cytometer.

The results from this experiment (Figure 3.10) suggest that plasmid size may have a marked effect on transfection efficiencies. The pmaxGFP plasmid is approximately a third of the size of pX458 (3.5kb vs. 9.3kb) and drives 6-fold greater expression of GFP. The same difference is seen when using a brighter variant of GFP. Cell densities following transfection appeared largely the same in all transfected wells suggesting this is not due to cytotoxicity. This must be taken into consideration when quantifying editing rates with the LoxP pX330 plasmids as their size is large (8.5kb) and smaller expression vectors are not readily available. A transfection, and hence editing, rate of ~20% is potentially what we would expect from a plasmid of this size.



C

ID	Plasmid	Size (Kb)	GFP (%)
1	N/A	N/A	0.1%
2	pmaxGFP	3.5	93.1%
3	pX458	9.3	15.3%
4	pX458-mWasabi	9.4	13.8%

Figure 3.10: Substantially less GFP expression is seen on using a larger expression vector in *mTmG* MEFs. A) Confocal images of each well taken at 20x 24 hours post-transfection. Scale bar 100µm. B) FC plots of cell count versus GFP intensity for each well at 28 hours post-transfection. C) Summary of transfection conditions for each well along with the resulting percentage of GFP+ cells.

3.5 Use of a Heterozygous *mTmG* Line Reduces but Does Not Eliminate the Double-Positive Population

Whilst we initially hypothesised that a homozygous *mTmG* cell line could report on editing at both alleles, extrapolating editing from the fluorescence observed, namely yellow or double positive, as well as fluorescence intensity (i.e. lower green intensity could be due several editing events) was overly complicated for our analyses. To monitor mono-allelic editing of the *mTmG* locus, the homozygous *Rosa26 mTmG* mouse line was crossed onto a CD1 line in order to generate heterozygous *mTmG* mice. MEFs were generated from these mice as before (Section 2.10.1) and were imaged on a Nikon A1R once they had reached confluency at P3. As before, all the cells could be seen to display membrane-Tomato and none membrane-EGFP. Importantly, there did not appear to be a significant decrease in fluorescence intensity (Figure 3.11).

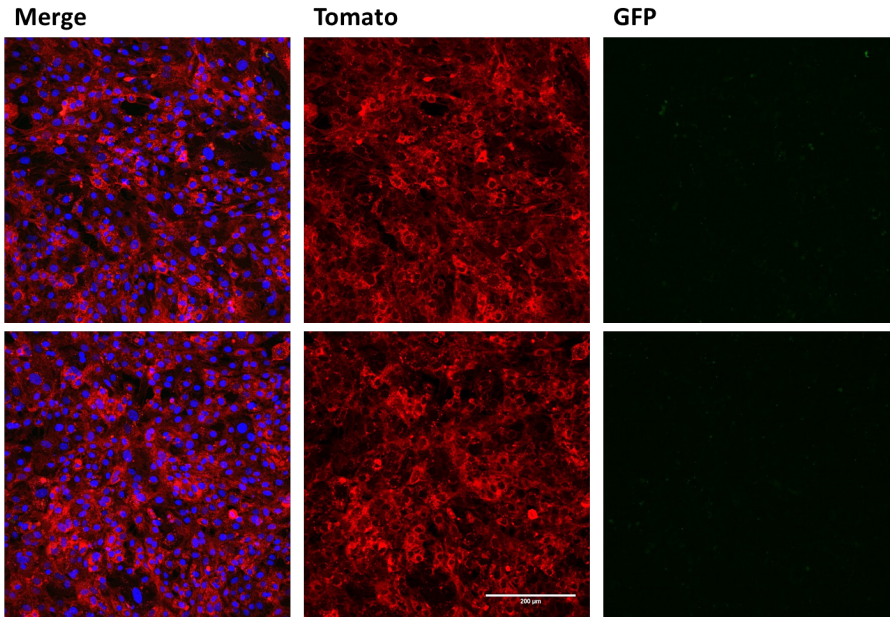


Figure 3.11: Heterozygous *mTmG* MEFs display Tomato fluorescence without a significant decrease in fluorescence intensity. Images captured at 20x. Scale bar 200µm.

To verify that the heterozygous MEFs also report accurately on genomic editing, they were transfected with LoxP#1 pX330 using the Neon System and 100µl tips (Section 2.12.1.1), $5x10^5$ cells per well. Seventy-two hours post-transfection, the cells were imaged on a Nikon A1R confocal. The cells were then dissociated and analysed by FC on a BD FACSAria seven days post-transfection. FC results (Figure 3.12) indicated

there is still a significant double-positive population present, although it is much reduced compared to in the homozygous MEFs, from ~60% of the non-red population (Figure 3.6) to ~11% (Here the non-red or 'edited' population is the total of the cells which fall in quadrants 1 to 3 (Figure 3.6), that is the total of the double-negative, GFP-positive and double-positive populations). The majority of edited cells are now found in the negative and GFP-positive quadrants (Q3 and Q4 respectively). This suggests that a large proportion of the double-positive cells were mono-allelically edited, however, another explanation must be found for the remaining double-positive cells. The long time between transfection and sorting (7 days), combined with the very minimal reduction in Tomato fluorescence intensity in this double-positive population (Q2 cf. Q1), suggests alternate possibilities to the half-life of the Tomato protein.

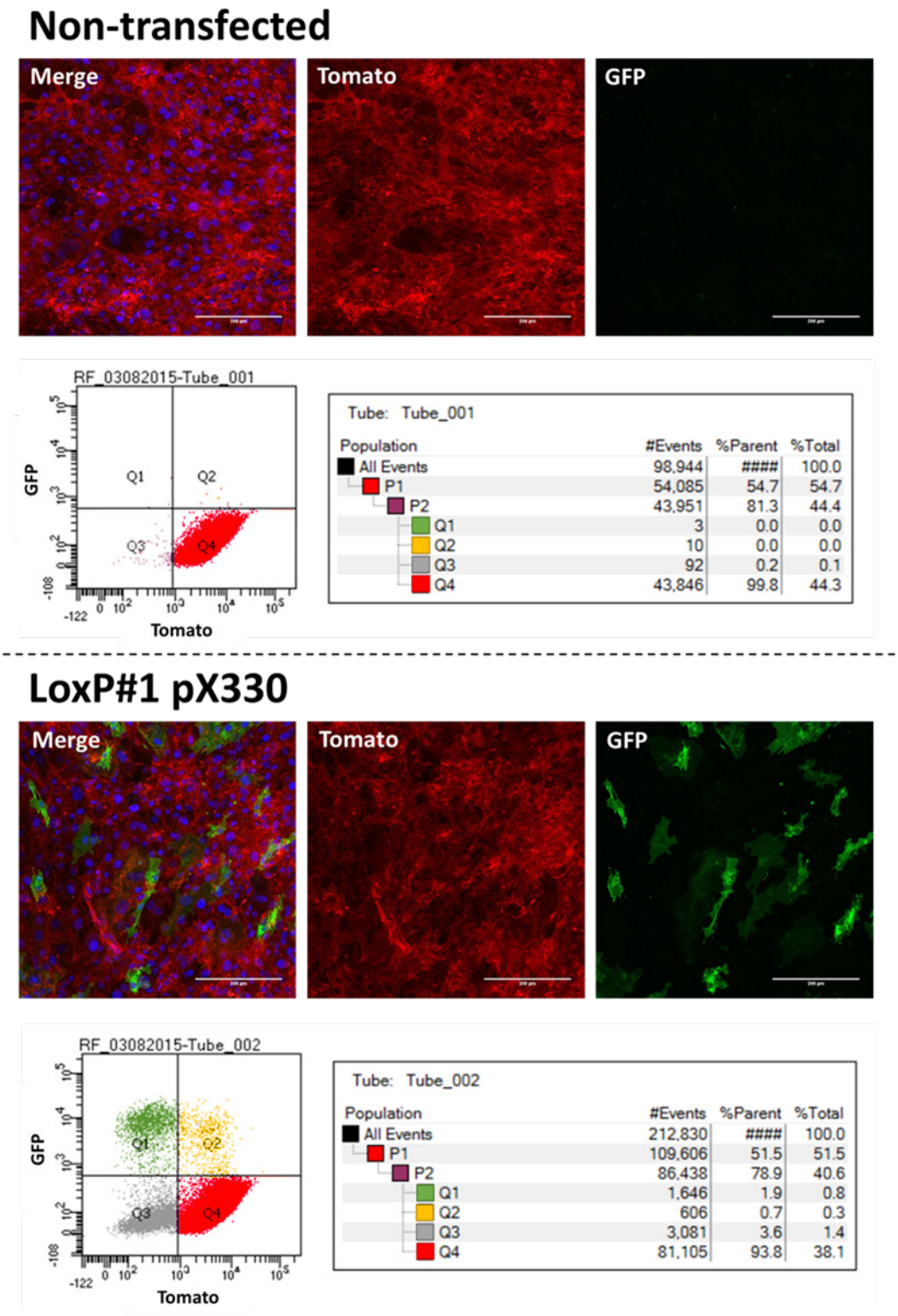


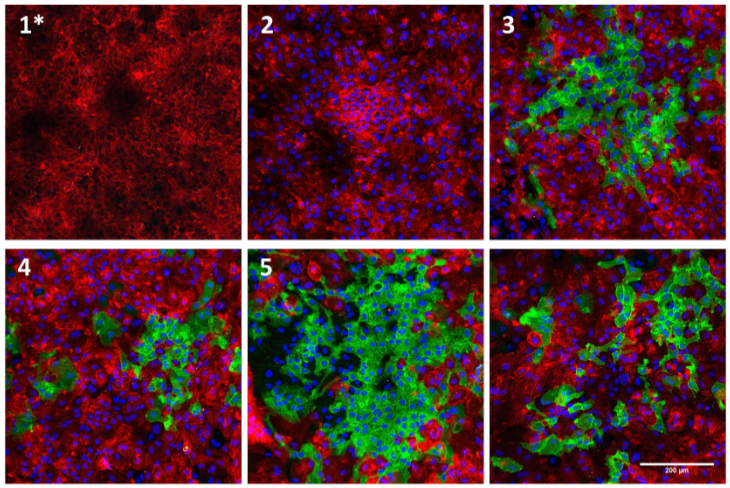
Figure 3.12: The size of the double-positive population (Q2) is substantially reduced in heterozygous *mTmG* MEFs compared to homozygous *mTmG* MEFs. Above, wells were imaged at 20x on a confocal microscope 72 hours post-transfection. Scale bar 200µm. Below, the results of an FC analysis of the wells carried out 7 days post-transfection.

3.6 Improving Rates of Editing in *mTmG* Ependymal Cells – Further Optimisation of the Neon Protocol

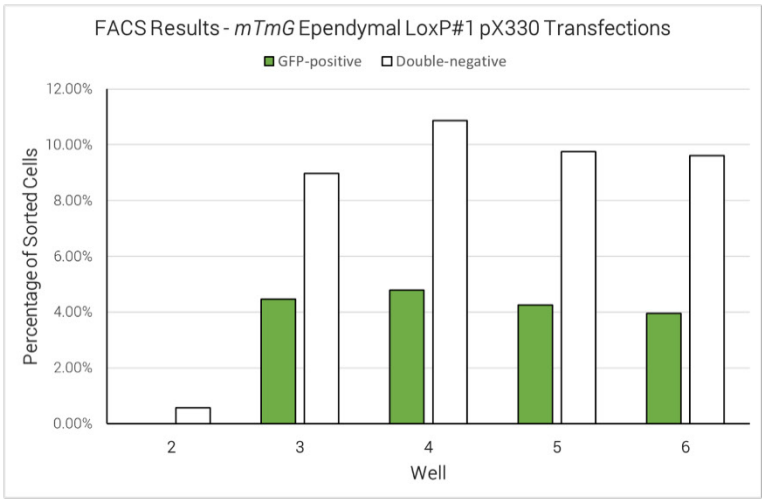
Given the reasonably low editing efficiencies seen in the initial Neon transfection experiment (Section 3.4.2), I decided to screen a range of conditions with the aim of increasing the transfection efficiency. Using heterozygous *mTmG* progenitor cells, 6 transfections were performed with the Neon 10µl tips as described in Section 2.12.1.2, except $2.4x10^5$ cells were transfected per well. Here the voltage and plasmid concentration were varied (Figure 3.13C). Following electroporation these cells were successfully differentiated into ependymal cells (Section 2.11.2). Each well was then imaged 8 days post-transfection on a Nikon A1R confocal microscope at 20x. Twenty-four hours after imaging, wells were dissociated using TrypLE Express (Thermo Fisher) at 37°C for thirty minutes. The long incubation time was required as these cells are strongly adherent. The dissociated cells were then re-suspended in PBS and analysed on a BD FACSAria.

Imaging and FC results indicated that there were much higher editing efficiencies in this experiment compared to previously; editing rates approaching 16% were seen (Figure 3.13B). Partially, this may be due to the increased sensitivity of FC as opposed to imaging, although it may also be due to batch-to-batch variation between *mTmG* progenitors. Changing the voltage or plasmid concentration within the range assayed did not appear to have a significant effect on rates of editing.

A



B



C

Well	Plasmid	DNA/reaction (ng)	Electroporation Parameters	Edited Cells (%)
1	N/A	0	1300V, 20ms, 2 pulses	N/A*
2	*Zmynd10* pX330	900	1300V, 20ms, 2 pulses	0.57%
3	LoxP#1 pX330	900	1250V, 20ms, 2 pulses	13.44%
4	LoxP#1 pX330	900	1300V, 20ms, 2 pulses	15.67%
5	LoxP#1 pX330	900	1350V, 20ms, 2 pulses	14.01%
6	LoxP#1 pX330	1200	1300V, 20ms, 2 pulses	13.57%

Figure 3.13: Voltage and plasmid concentration do not have a large effect on ependymal Neon transfection efficiencies in the ranges assayed. A) 20x confocal images of the wells were taken at 8 days post-transfection, following differentiation. Generally, highly clonal populations of edited cells could be seen. Scale bar 200µm. B) Cells were analysed by FC 9 days post-transfection and the results graphically plotted. FC plots included on attached CD in file 'Figure 3_13'. C) Summary of transfection conditions for each well along with the resulting rate of total editing. Note: well 1 was imaged without a nuclear stain and was also not analysed by FC. Well 2 was used as a negative FC control.

3.7 Membrane-tagged Fluorescent Reporters Allow Easy Identification of Edited Cell Types

The ependymal cells from the preceding section were imaged at higher magnifications; 40x and 80x using oil immersion on a Nikon A1R confocal microscope, 24 hours prior to being analysed by FC. These images demonstrate that the membrane-tagged fluorescent reporters allow simple visual identification of the edited cell types based on morphology and motility; the ependymal cells can clearly be identified based on their size, shape and formation of tight junctions. Furthermore, both membrane-Tomato and membrane-EGFP can be seen to be present in motile cilia, indicating we have achieved editing without affecting the differentiation of motile ciliated cells (Figure 3.14).

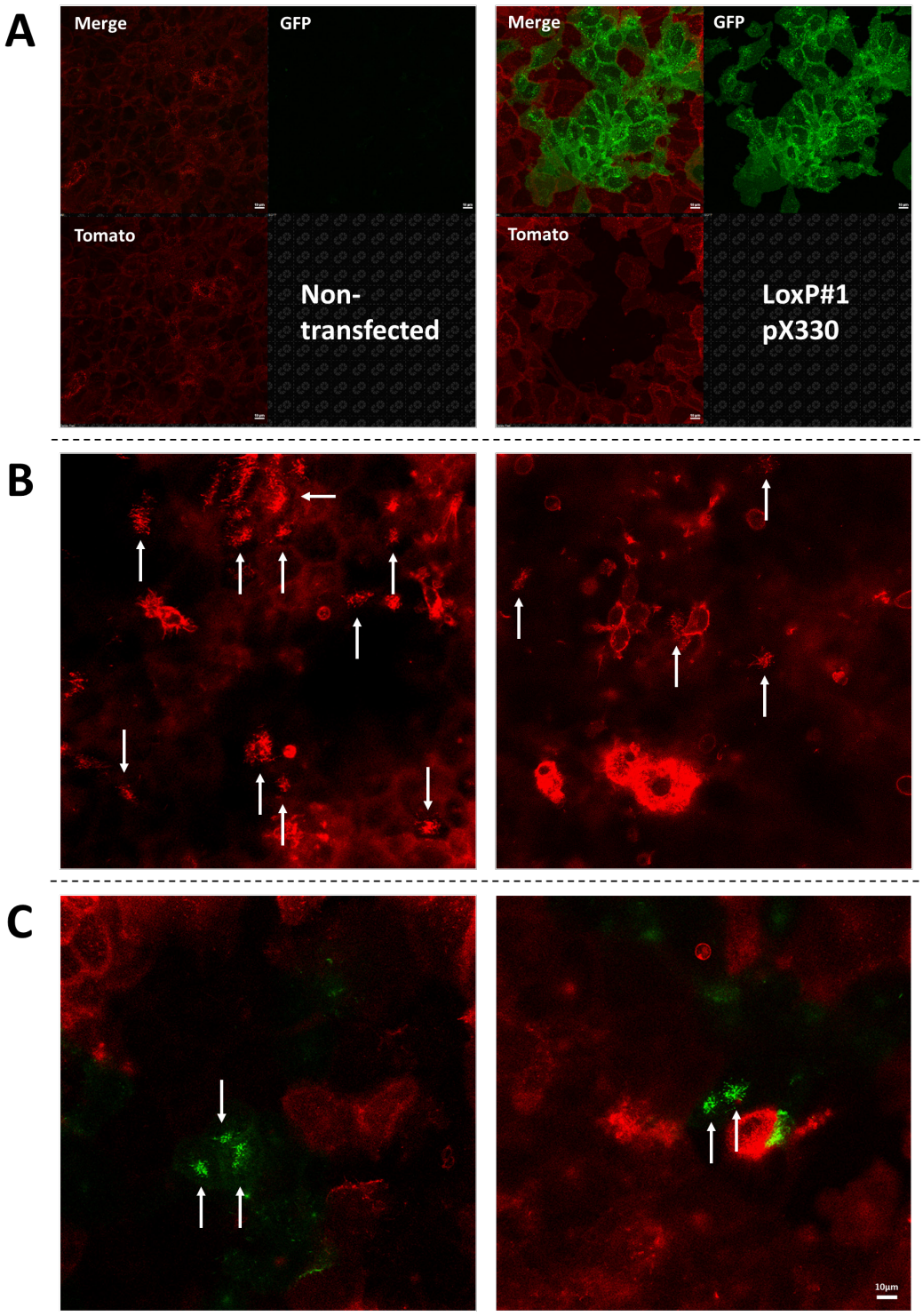


Figure 3.14: The membrane-tagged fluorescent reporters allow identification of the edited cell types (here: ependymal cells) based on morphology. A) 40x confocal images of the ependymal cells reveal that cell morphology is clearly delineated by the membrane-tagged reporters. B) 80x confocal images of the non-transfected sample reveal that membrane-Tomato is visible in the motile cilia. C) 80x confocal images of the LoxP#1 pX330 transfected sample reveal that membrane-EGFP is also visible in the motile cilia. White arrows: patches of multi-cilia. Scale bars 10µm

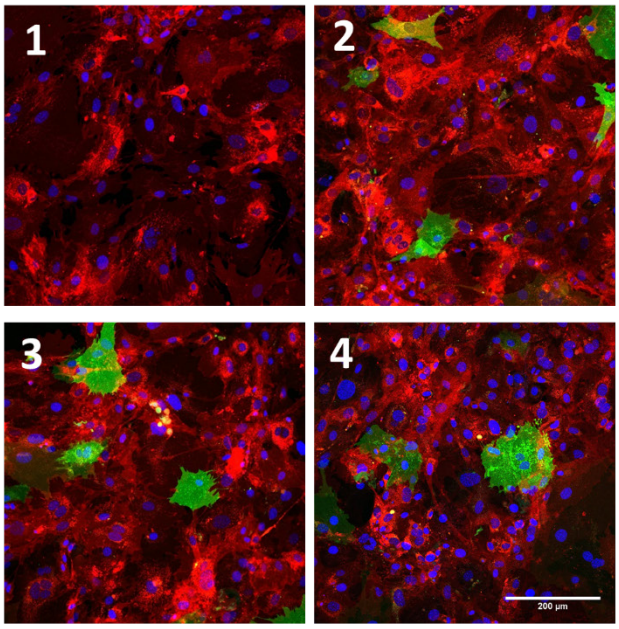
3.8 Combining Guides against LoxP may Modestly Improve Rates of Editing

As mentioned previously, two guides were designed against the loxP locus (Section 3.2.1). The second guide was designed subsequent to the first guide, in order to ascertain whether our guide choice was limiting editing efficiencies. We also hypothesised that by targeting multiple loci simultaneously we may be able to improve rates of editing. To test this, I performed four transfections in the heterozygous *mTmG* MEFs; a no DNA control, LoxP#1 pX330, LoxP#2 pX330 and LoxP#1 pX330 + LoxP#2 pX330. The Neon System and the 100µl electroporation tips were used as described in Section 2.12.1.1 with $5x10^5$ cells per reaction. The total amount of DNA was kept the same for all transfections (that is 5000ng total per well). The transfected wells were imaged 4 days post-transfection on a Nikon A1R confocal microscope, and then analysed by FC the following day on a BD FACSAria. Both guides when used alone drove similar rates of editing, suggesting no significant sequence-based bias at this locus. Combining both guides may increase rates of editing, although only modestly (Figure 3.15C). Unexpectedly, in this experiment there was again a large proportion of double-positive cells (approximately 40% of non-red). Furthermore, a greater proportion of the non-transfected MEFs appeared to be losing fluorescence. This is likely a technical issue, such as a blockage or contamination during the FC analysis, as few single, live cells were detected (only ~6,000 vs. ~50,000 for wells 2-4, data not shown).

A

Well	Cell Number	Plasmid 1	Quantity (ng)	Plasmid 2	Quantity (ng)
1	$5x10^5$	N/A	0	N/A	0
2	$5x10^5$	LoxP#1 pX330	5000	N/A	0
3	$5x10^5$	LoxP#2 pX330	5000	N/A	0
4	$5x10^5$	LoxP#1 pX330	2500	LoxP#2 pX330	2500

B



C

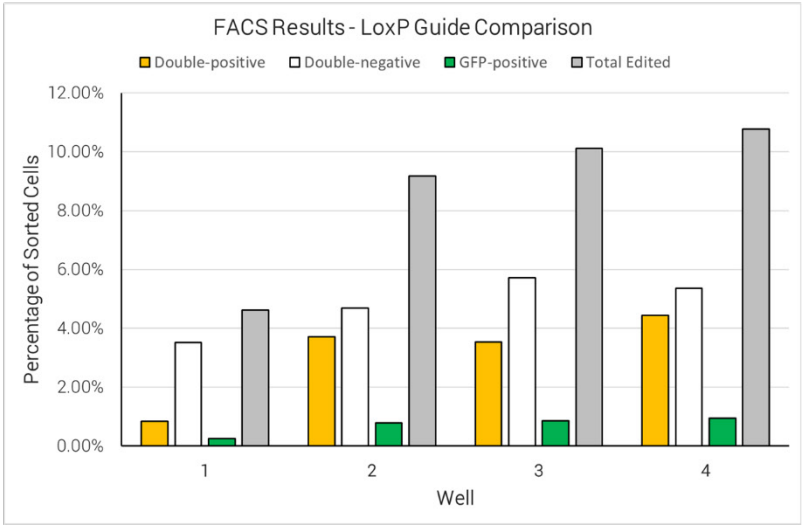


Figure 3.15: Both guides (independently) targeting the loxP sites generate similar levels of editing in *mTmG* MEFs. Combining the two guides appeared to have little effect. A) Summary of experimental set-up. B) Confocal images of each well captured at 20x, 96 hours post-transfection. Scale bar 200µm. C) Each well was analysed by FC 120 hours post-transfection and the results were plotted graphically. FC plots included on attached CD in file 'Figure 3_15'.

3.9 ssODNs Are Able to Drive HDR at Our Locus, Albeit at a Very Low Frequency

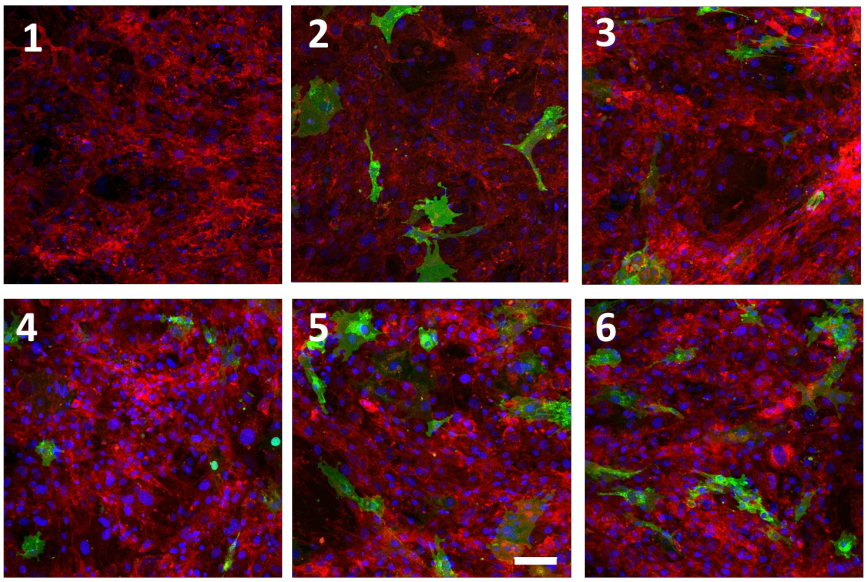
As explained in Section 3.2.2, three ~200nt ssODN repair templates were designed to remove the MARCKS membrane tag and affix a nuclear localisation signal to the EGFP in the *mTmG* mice. The three ssODN repair templates are termed NLS EGFP 1, NLS EGFP 2 and NLS EGFP 3, and are of varying designs (Figure 3.4). All three repair templates were synthesised by IDT. The three repair templates were transfected into heterozygous *mTmG* MEFs, together with LoxP#2 pX330, using the 100µl Neon electroporation tips as described in Section 2.12.1.1. Seventy-two hours post-transfection each well was imaged on a Nikon A1R confocal microscope. Seven days post-transfection, the wells were then analysed by FC on a BD FACSAria.

Confocal imaging revealed that the ssODN repair templates were very inefficient at driving HDR in this system. Wells were scanned manually for nuclear GFP and only two positive cells were observed (in a confluent 6 well plate); these were transfected with LoxP#2 pX330 and NLS EGFP 1. Promisingly, the GFP signal overlapped perfectly with the Hoechst signal. These cells also appeared to be Tomato-negative indicating correct repair (Figure 3.16C). FC results again revealed similar editing rates with either LoxP#1 pX330 or LoxP#2 pX330 (FC does not distinguish nuclear- from membrane-GFP, see Section 3.14). Interestingly, rates of editing appeared to be slightly higher when a repair template was concurrently supplied (Figure 3.16D).

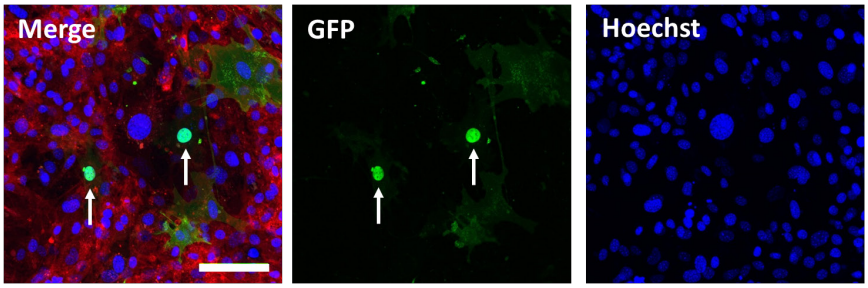
A

Well	Cell Number	Cas9 Plasmid	Quantity (ng)	ssODN	Quantity
1	$5x10^5$	N/A	0	N/A	0
2	$5x10^5$	LoxP#1 pX330	5000	N/A	0
3	$5x10^5$	LoxP#2 pX330	5000	N/A	0
4	$5x10^5$	LoxP#2 pX330	5000	NLS EGFP 1	10µM, 2.5µl
5	$5x10^5$	LoxP#2 pX330	5000	NLS EGFP 2	10µM, 2.5µl
6	$5x10^5$	LoxP#2 pX330	5000	NLS EGFP 3	10µM, 2.5µl

B



C



D

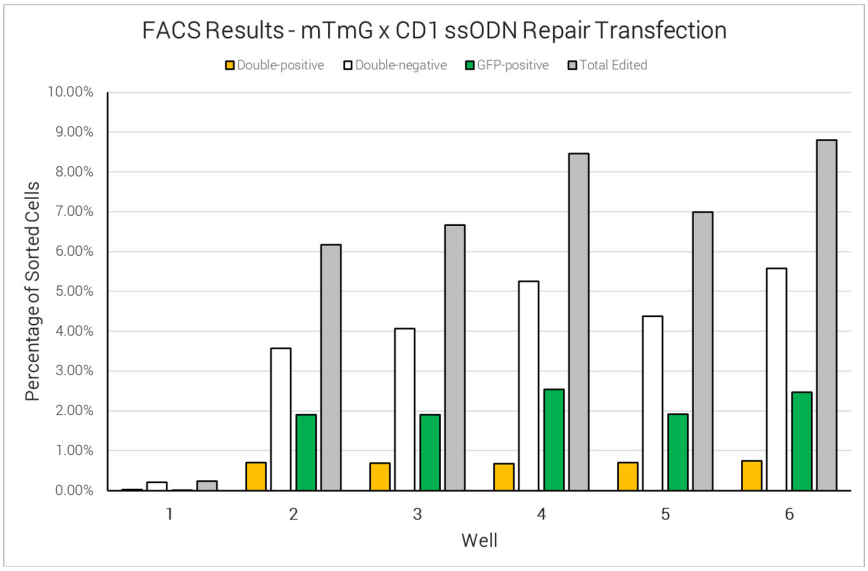


Figure 3.16 (on previous page): ssODN repair templates are inefficient at driving HDR in the *mTmG* MEFs. A) Summary of experimental set-up. B) Confocal images were taken of each well at 20x, 72 hours post-transfection. Scale bar 100µm. C) 40x image of the cells with nuclear-GFP, which also appear to be Tomato-negative. Scale bar 100µm. White arrows: nuclear-GFP. D) 7 days post-transfection cells were analysed by FC and the results were plotted graphically. FC plots included on attached CD in file 'Figure 3_16'.

3.10 Increasing Rates of Repair by Increasing Homology – Design of *mTmG* Repair Plasmids

Whilst ssODN repair templates are very effective for introducing small changes by HDR (Ran, Hsu et al. 2013), editing in our reporter system is on the scale of several kilobases and the ~60nt homology arms may not be sufficient. In order to provide significantly larger regions of homology, we engineered 2 plasmids each containing ~800bp homology arms extending into the pCA promoter at the 5' end and the EGFP at the 3' end. Each has both loxP PAM sites mutated to prevent re-cleavage. The designs of the two original repair plasmids, NLS Repair and H2B Repair, are shown in Figure 3.4.

Both plasmids were transfected into heterozygous *mTmG* MEFs alongside LoxP#1 pX330 or LoxP#2 pX330 using the Neon 100µl electroporation tips (Section 2.12.1.1). Furthermore, one well was prepared to test the effect of doubling the total plasmid concentration (Figure 3.17A). Seventy-two hours post-transfection wells were imaged in ten random positions at 20x on a Nikon A1R confocal. The resulting images were then quantified using the Nucleus Counter and Cell Counter plugins in Fiji to estimate the percentage of cells which have undergone NHEJ and HDR.

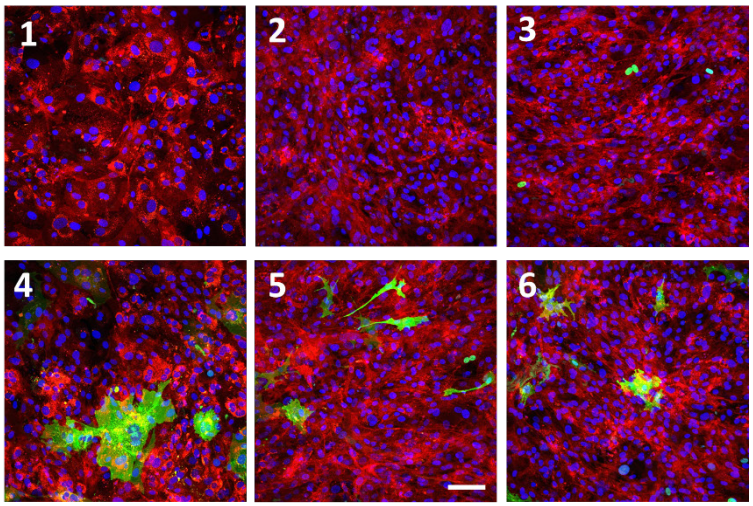
Unexpectedly, the images revealed that both plasmids, especially the H2B repair plasmid, are able to drive expression of nuclear-GFP in the absence of editing and/or repair (Figure 3.17). It is unclear how this would arise at this frequency as the repair plasmids are promoterless; their 780bp 5' homology arms contain only part of the Chicken β-Actin intron and the Rabbit β-Globin splice acceptor. However, these results did indicate that the histone H2B sequence results in brighter nuclear-GFP than the SV40 NLS sequence. It is unclear why there appeared to be ~6 fold more edited cells in well 4 with LoxP#1, as previously we had seen similar results with both plasmids. The results from this experiment also indicated that doubling the total

plasmid concentration (by adding an equal amount of repair plasmid) appears to be well tolerated and offers greater rates of editing (cf. well 5 and 6). Based on this, in all following experiments 1000ng of Cas9 plasmid(s) and 1000ng of repair plasmid (2000ng total) were used per $1x10^5$ cells.

A

Well	Cell Number	Cas9 Plasmid	Quantity (ng)	Repair Plasmid	Quantity (ng)
1	$5x10^5$	N/A	0	N/A	0
2	$5x10^5$	N/A	0	NLS Repair	2500
3	$5x10^5$	N/A	0	H2B Repair	2500
4	$5x10^5$	LoxP#1 pX330	2500	NLS Repair	2500
5	$5x10^5$	LoxP#2 pX330	2500	H2B Repair	2500
6	$5x10^5$	LoxP#2 pX330	5000	H2B Repair	5000

B



C

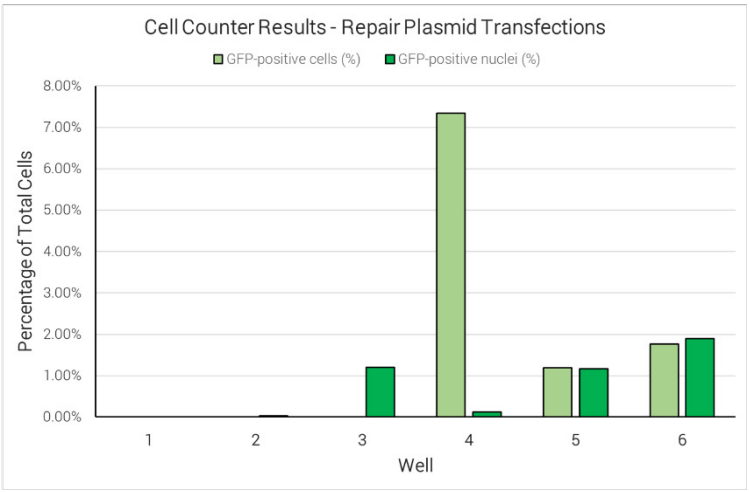


Figure 3.17: Promoterless repair plasmids drive expression of nuclear-GFP in the absence of editing in *mTmG* MEFs. A) Summary of experimental set-up. B) Representative confocal images of each well taken at 20x 72-hours post-transfection. Scale bar 100µm. C) Confocal images were quantified using Nucleus Counter and Cell Counter plugins in Fiji and the results were plotted graphically.

3.11 Creation of Immortalised *mTmG* MEF Lines to Facilitate Experiments

MEFs have a limited life-span and can only be propagated for approximately 8 passages before becoming senescent. This significantly hinders our ability to expand cells following FACS for imaging and/or to extract a practical quantity of DNA for sequencing. In order to circumvent this problem, I immortalised the heterozygous *mTmG* MEFs with an SV40 large T antigen plasmid as described in Section 2.10.4. Following immortalisation, the MEFs appeared morphologically normal, although their doubling time was significantly reduced. The immortalised heterozygous *mTmG* MEFs were used in all following experiments, unless otherwise noted.

3.12 Re-designing the H2B Repair Plasmid to Limit Aberrant Expression

In order to attempt to eliminate transient expression, a new repair plasmid, termed H2Bv2, was designed in which the final 10 codons (including stop) were removed from EGFP, whilst maintaining ~740bp of 3' homology. Without the final 10 codons, we hypothesised that the EGFP would be unstable in the absence of HDR. I transfected this new repair plasmid into the immortalised heterozygous *mTmG* MEFs alongside LoxP#1 and/or LoxP#2 pX330 using the Neon 100µl tips:

Well	Cell Number	Cas9 Plasmid	Quantity (ng)	Repair Plasmid	Quantity (ng)
1	$5x10^5$	N/A	0	N/A	0
2	$5x10^5$	N/A	0	H2Bv2	5000
3	$5x10^5$	LoxP#1 pX330	5000	H2Bv2	5000
4	$5x10^5$	LoxP#2 pX330	5000	H2Bv2	5000
5	$5x10^5$	LoxP#1 pX330	5000	H2Bv2	5000

Forty-eight hours post-transfection the wells were imaged on a Nikon A1R confocal microscope at 40x. Unexpectedly, these images indicated that the new repair plasmid could still drive transient expression of nuclear-GFP (Figure 3.18A). To gauge the duration of the transient plasmid expression, I decided to maintain the cells further. The cells were imaged again at 144 hours post-transfection, after having been split 1-in-2 twenty-four hours previously. At this time-point no nuclear-GFP signal could be seen in the H2Bv2 repair template alone well, whereas it was maintained in the

pX330 wells indicating repair. Critically, the cells with nuclear-GFP appeared to have lost Tomato fluorescence (Figure 3.18C). The results revealed HDR in approximately 0.5% of total cells, when combining both LoxP pX330 plasmids (Figure 3.18D).

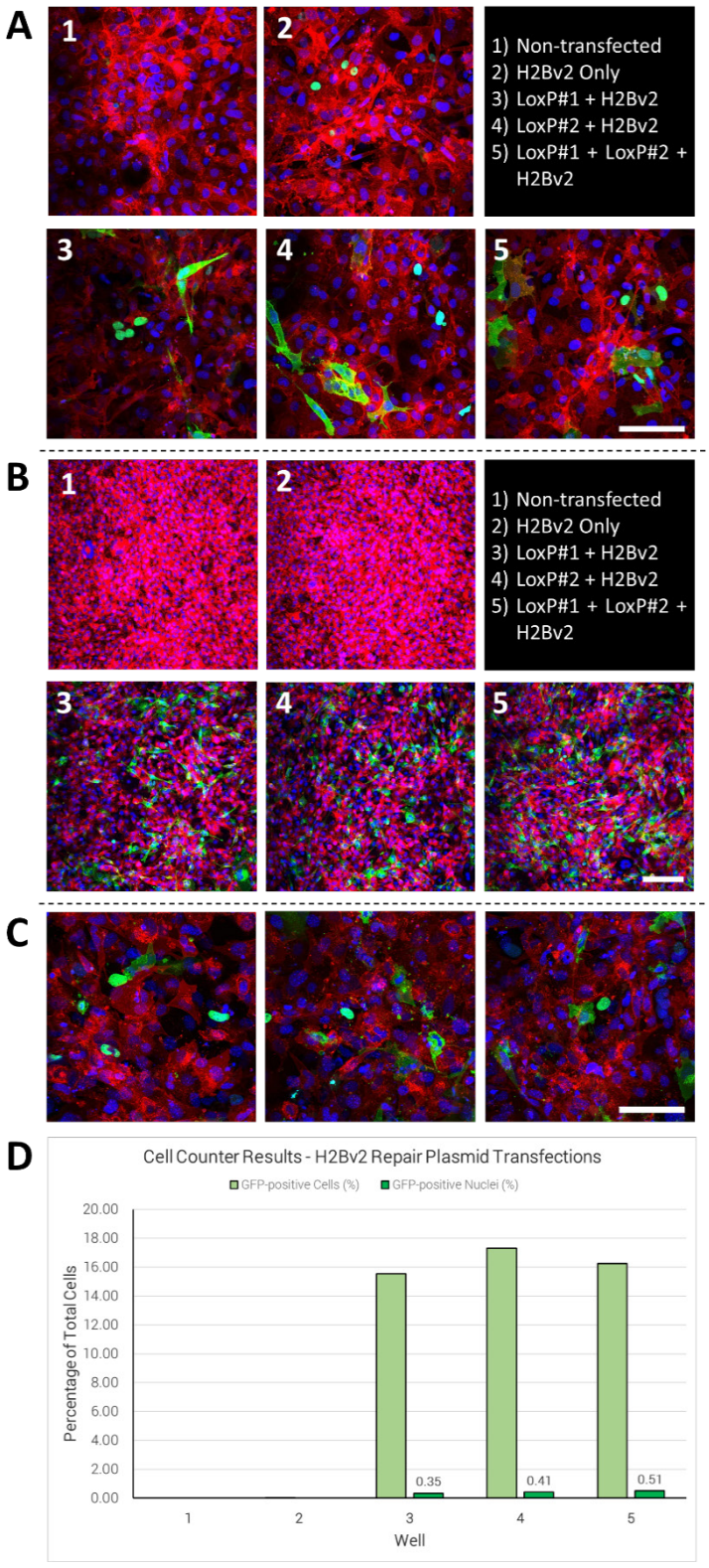


Figure 3.18 (on previous page): Plasmid-based nuclear-GFP expression from the H2Bv2 construct disappears after 144 hours in MEFs. A) 40x confocal images at 48 hours post-transfection. Scale bar 100µm. Top right: well layout. B) 15x confocal images at 144 hours post-transfection. Scale bar 250µm. Top right: well layout. C) 40x close up of the GFP-positive nuclei seen in well 5. Scale bar 100µm. D) Cell Counter and Nucleus Counter quantification of the images taken at 144 hours post-transfection.

3.13 Membrane- and Nuclear-EGFP are Not Distinguishable by FACS

Given our confocal imaging capabilities are relatively low throughput, sampling only a small proportion of the cells in the well, it was hoped we could develop a FACS-based approach for counting cells with nuclear-GFP. We had envisioned that cells with nuclear-GFP would have a lower fluorescence intensity than cells with membrane-GFP and could be counted and isolated as a sub-population on a FACS plot. To test this directly, I performed a transfection with LoxP#1 pX330, LoxP#2 pX330 and H2Bv2 as before, along with non-transfected and H2Bv2 only controls. These cells were maintained for 14 days prior to being sorted on a BD FACSAria, by which point no transient nuclear-GFP fluorescence could be seen in the H2Bv2 only control.

The transfected cells then were sorted into five populations: Tomato-positive, double-positive, bright GFP (i.e. membrane), dim GFP (i.e. nuclear) and double-negative. The sorted cells were then returned to glass-bottom plates and allowed to reach confluency. At this point, each well was imaged on a Nikon A1R confocal. Imaging revealed that both dim and bright GFP populations contained many cells with GFP-positive nuclei, as well as cells with membrane-GFP (Figure 3.19B), however, the dim GFP population appeared to be more similar to the double-positive population. This may be reflective of editing at the sequence level partially disrupting varying degrees of the surrounding sequence, such as the promoter, and therefore changing expression levels. Imaging also revealed that there was significant contamination from the Tomato-positive cells, which are the most prevalent, into all the sorted populations. These results suggest that our current FACS system (BD FACSAria) is unable to reliably separate and/or count the nuclear-GFP population by fluorescence intensity.

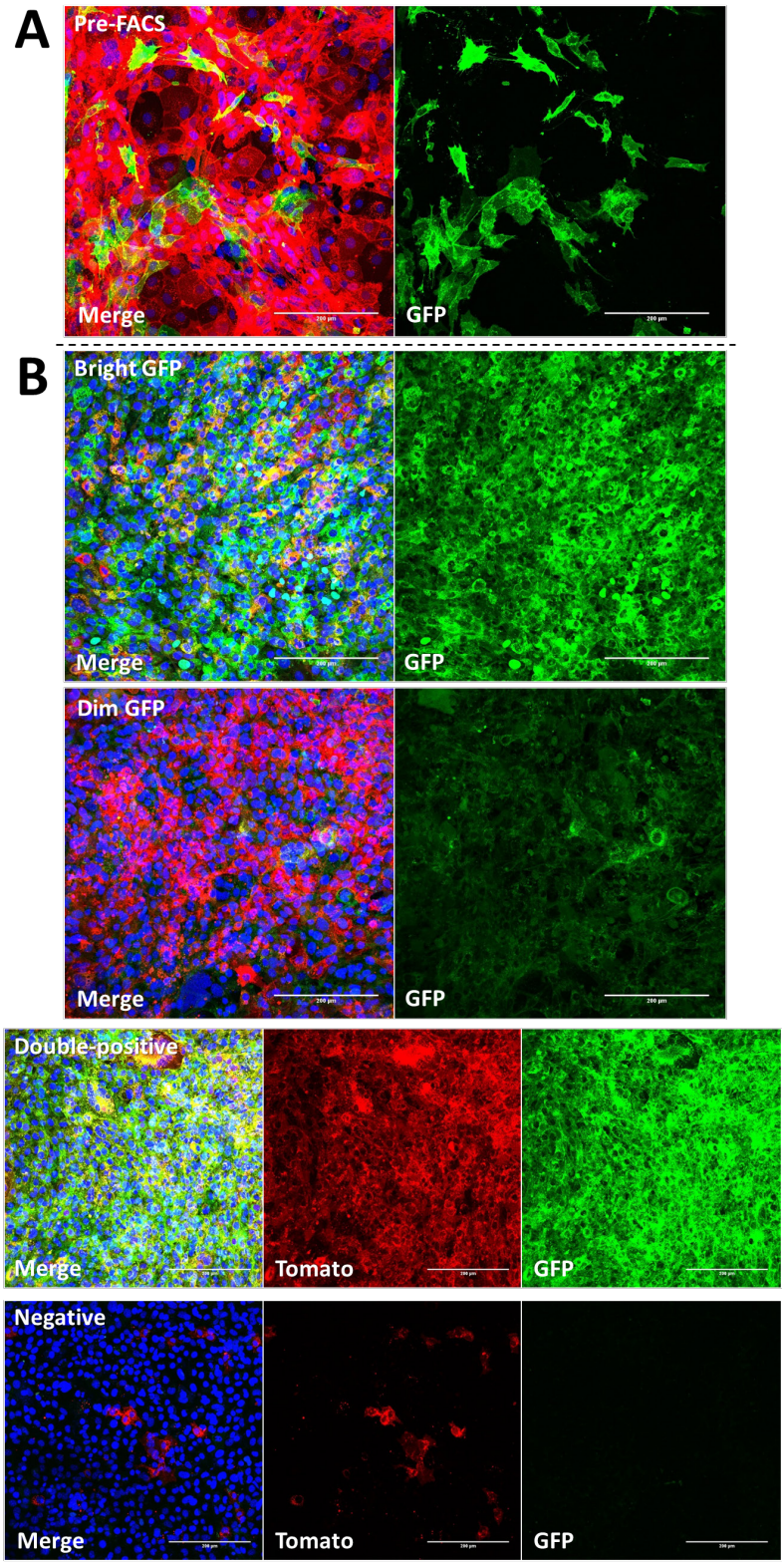


Figure 3.19: FACS cannot reliably distinguish nuclear-GFP from membrane-GFP in *mTmG* MEFs. A) Confocal image of the LoxP + H2Bv2 well captured at 20x prior to sorting by FACS. B) Confocal images of sorted populations captured at 20x. Scale bar 200µm.

3.14 Modulating Double-Stranded DNA Break Repair with Reported Enhancer Small Molecules BFA, L-755,507, RS-1 at the mTmG Locus

Given our system provides a visual read-out of repair, we hypothesised it would be amenable to small molecule screening protocols; with the ultimate goal of identifying novel enhancers of HDR. In a small proof-of-principle experiment, I tested the effect of four small molecules which have been shown to improve rates of HDR in our system. Three separate experiments were carried out, varying different parameters, in order to identify the optimum concentration and time-point to use each molecule. In the first experiment (Figure 3.20), three concentrations of BFA, L-755,507 and SCR7 were assayed. In the second (Figure 3.21), different time points for the addition of the same small molecules were tested. Finally, based on the most promising results from the previous two experiments, in the third (Figure 3.22), further concentrations of BFA and SCR7 were evaluated, along with five concentrations of another small molecule, RS-1. All small molecules were removed 24 hours after addition. The transfections were all performed using Neon 10µl tips and a combination of LoxP#1 pX330, LoxP#2 pX330 and H2Bv2, as this had been seen previously to give the greatest rate of HDR:

Cell Number	Plasmid 1	Quantity (ng)	Plasmid 2	Quantity (ng)	Plasmid 3	Quantity (ng)
$0.5x10^5$	LoxP#1 pX330	250	LoxP#2 pX330	250	H2Bv2	500

Five days post-transfection, the wells were split 1 in 2 to disrupt clonal populations and allowed to settle for 6 hours. Each well was then imaged 10 times at 20x in random positions on a Nikon A1R confocal microscope. Images were quantified using Cell Counter and Nucleus Counter (Section 2.16.1). Following imaging, plate 3 was also analysed on a BD LSRFortessa to determine total rates of editing.

Generally, there appeared to be little effect on the rates of NHEJ seen with any of the treatments (Figure 3.22C), whereas the rates of HDR picked up by imaging were highly variable with no clear trend for any of the small molecules (Figure 3.22B). As none of the conditions assayed resulted in a significant improvement in the rates of

HDR over the control well (well 3A, ~0.5% HDR), while the current methodology employed was highly time-consuming and gave variable results, it was decided to not continue with further screening using this approach. Alternative platforms, which may provide more consistent results when detecting HDR in our system, were not available on-site at this time.

A

Well	Plasmid(s)	Compound	Concentration (µM)	Volume of Drug (µl)	Additional DMSO (µl)	Time Added
1A	No DNA	N/A	N/A	0.000	5.000	t+0:00
2A	H2Bv2	N/A	N/A	0.000	5.000	t+0:00
3A	LoxP#1 + #2 + H2Bv2	N/A	N/A	0.000	5.000	t+0:00
4A	LoxP#1 + #2 + H2Bv2	BFA	0.0500	1.000	4.000	t+0:00
5A	LoxP#1 + #2 + H2Bv2	BFA	0.0250	0.500	4.500	t+0:00
6A	LoxP#1 + #2 + H2Bv2	BFA	0.0125	0.250	4.750	t+0:00
7A	LoxP#1 + #2 + H2Bv2	SCR7	0.2000	2.000	3.000	t+0:00
8A	LoxP#1 + #2 + H2Bv2	SCR7	0.0500	0.500	4.500	t+0:00
9A	LoxP#1 + #2 + H2Bv2	SCR7	0.0250	0.250	4.750	t+0:00
10A	LoxP#1 + #2 + H2Bv2	L-755,507	10.0000	5.000	0.000	t+0:00
11A	LoxP#1 + #2 + H2Bv2	L-755,507	2.5000	1.250	3.750	t+0:00
12A	LoxP#1 + #2 + H2Bv2	L-755,507	1.2500	0.625	4.375	t+0:00

B

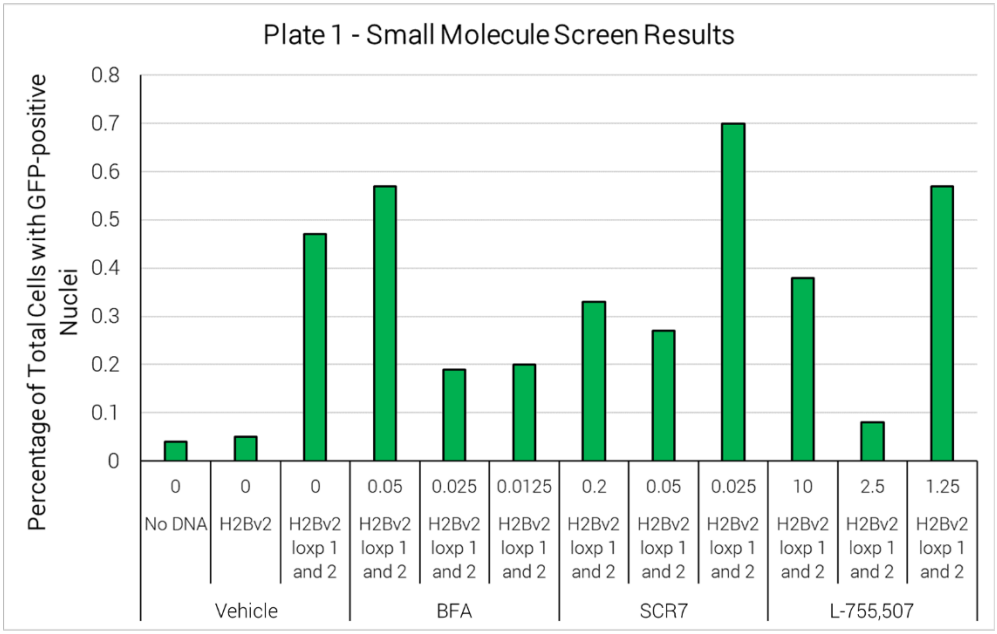


Figure 3.20: The small molecules BFA, SCR7 and L-755,507 do not appear to have a significant effect on HDR at any concentration tested in *mTmG* MEFs. A) Summary of experimental set-up. B) Ten 20x images were taken of each well in random positions and then quantified using the Nucleus Counter and Cell Counter plugins in Fiji.

A

Well	Plasmid(s)	Compound	Concentration (µM)	Volume of Drug (µl)	Additional DMSO (µl)	Time Added
1B	No DNA	N/A	N/A	0	2.5	t+00:00
2B	No DNA	N/A	N/A	0	2.5	t+04:00
3B	No DNA	N/A	N/A	0	2.5	t+24:00
4B	LoxP#1 + #2 + H2Bv2	BFA	0.1	2	0.5	t+00:00
5B	LoxP#1 + #2 + H2Bv2	BFA	0.1	2	0.5	t+04:00
6B	LoxP#1 + #2 + H2Bv2	BFA	0.1	2	0.5	t+24:00
7B	LoxP#1 + #2 + H2Bv2	SCR7	0.1	1	1.5	t+00:00
8B	LoxP#1 + #2 + H2Bv2	SCR7	0.1	1	1.5	t+04:00
9B	LoxP#1 + #2 + H2Bv2	SCR7	0.1	1	1.5	t+24:00
10B	LoxP#1 + #2 + H2Bv2	L-755,507	5	2.5	0	t+00:00
11B	LoxP#1 + #2 + H2Bv2	L-755,507	5	2.5	0	t+04:00
12B	LoxP#1 + #2 + H2Bv2	L-755,507	5	2.5	0	t+24:00

B

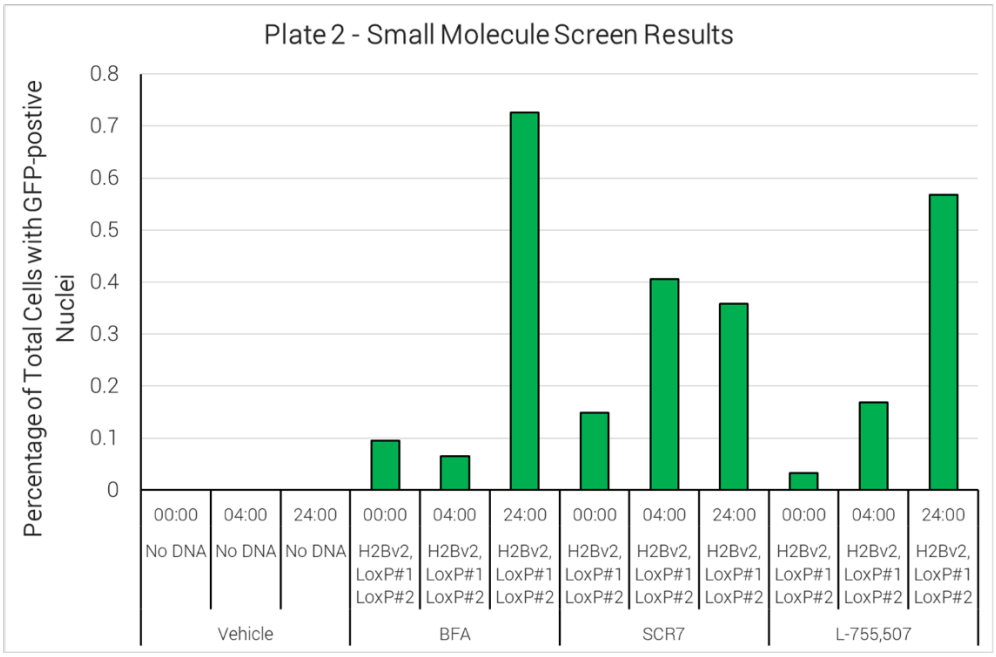
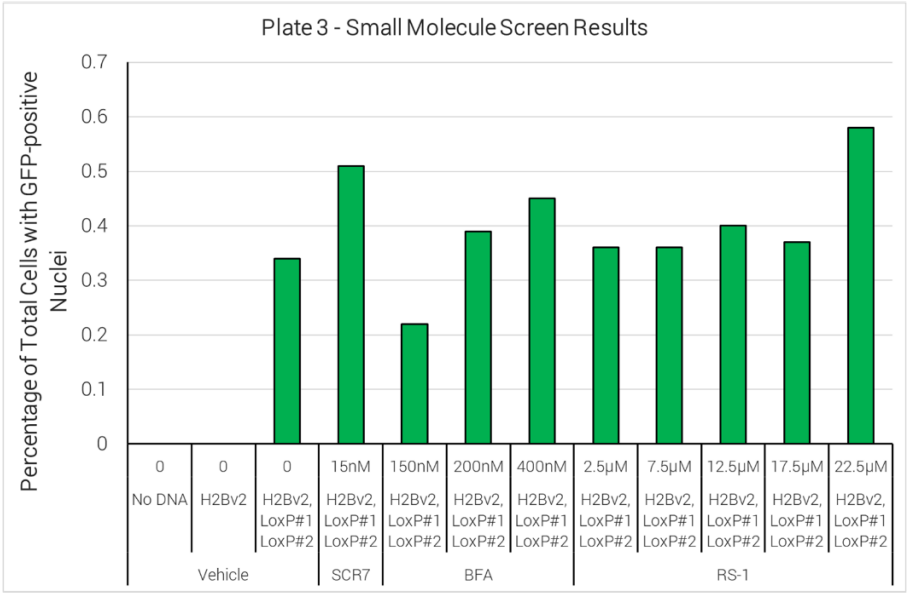


Figure 3.21: The highest rates of HDR appear to be achieved by adding the small molecules 24 hours post-transfection in *mTmG* MEFs. A) Summary of experimental set-up. B) Ten 20x images were taken of each well at random and then quantified using the Nucleus Counter and Cell Counter plugins in Fiji.

A

Well	Plasmid(s)	Compound	Concentration (µM)	Volume of Drug (µl)	Additional DMSO (µl)	Time Added
1C	No DNA	N/A	N/A	0.000	11.250	t+24:00
2C	H2Bv2	N/A	N/A	0.000	11.250	t+24:00
3C	LoxP#1 + #2 + H2Bv2	N/A	N/A	0.000	11.250	t+24:00
4C	LoxP#1 + #2 + H2Bv2	SCR7	0.0150	0.150	11.100	t+24:00
5C	LoxP#1 + #2 + H2Bv2	BFA	0.1500	3.000	8.250	t+24:00
6C	LoxP#1 + #2 + H2Bv2	BFA	0.2000	4.000	7.250	t+24:00
7C	LoxP#1 + #2 + H2Bv2	BFA	0.4000	8.000	3.250	t+24:00
8C	LoxP#1 + #2 + H2Bv2	RS-1	2.5000	1.250	10.000	t+24:00
9C	LoxP#1 + #2 + H2Bv2	RS-1	7.5000	3.750	7.500	t+24:00
10C	LoxP#1 + #2 + H2Bv2	RS-1	12.5000	6.250	5.000	t+24:00
11C	LoxP#1 + #2 + H2Bv2	RS-1	17.5000	8.750	2.500	t+24:00
12C	LoxP#1 + #2 + H2Bv2	RS-1	22.5000	11.250	0.000	t+24:00

B



C

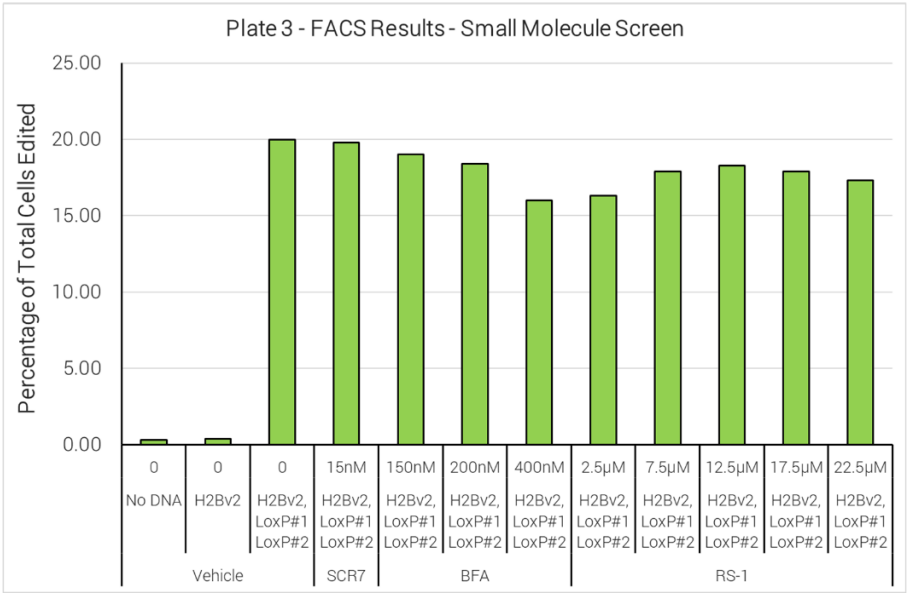


Figure 3.22: The small molecules BFA, SCR7 and RS-1 do not appear to have an effect on HDR or NHEJ at any concentration tested in *mTmG* MEFs. A) Summary of experimental set-up. B) Ten 20x images were taken of each well at random and then quantified using the Nucleus Counter and Cell Counter plugins in Fiji. C) Following imaging, each well was analysed by FC to quantify the total rate of editing. FC plots included on attached CD in file 'Figure 3_22'.

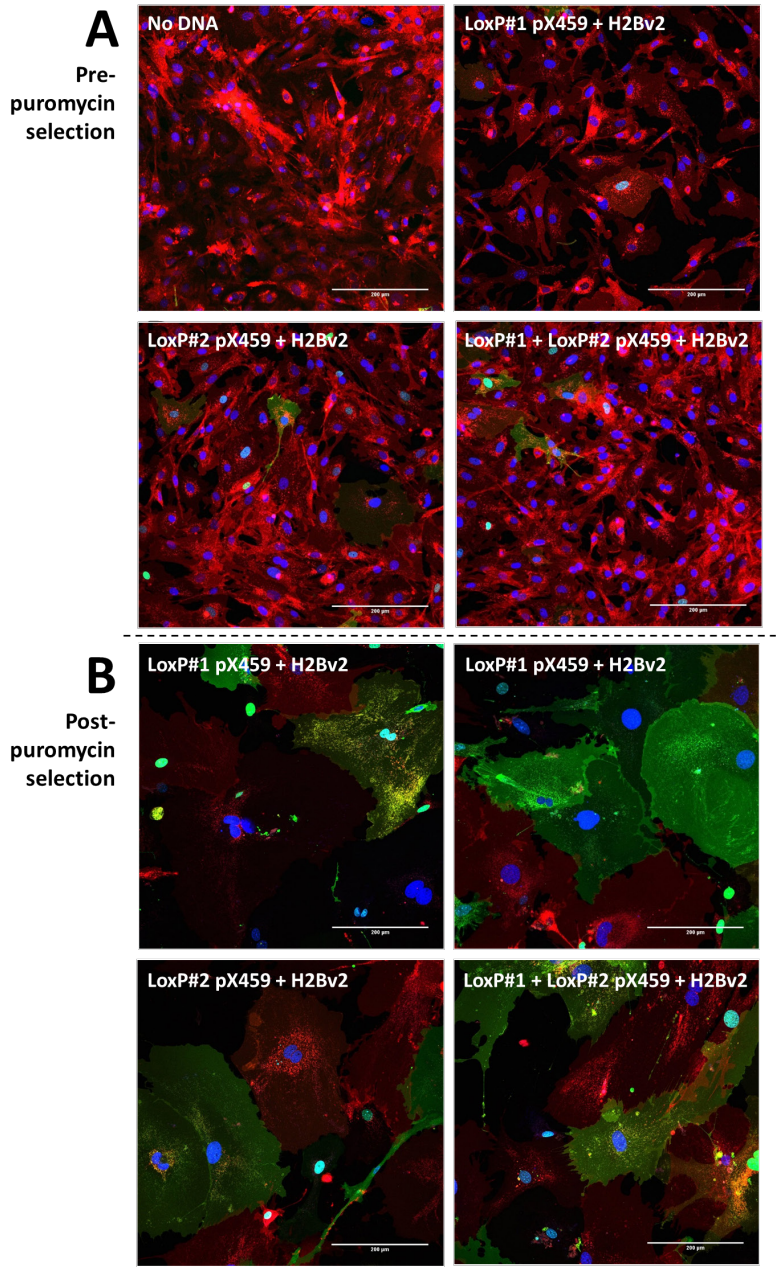
3.15 Efficiency of Editing versus Efficiency of Transfection – Puromycin Selection Reveals the Majority of Transfected Cells are Edited

Given the large discrepancy between the number of GFP-positive cells following pmaxGFP transfection and LoxP pX330 transfection, we needed to examine the relationship between the transfection rate and the editing rate in our system. In order to achieve this, I cloned both guides (Section 2.3.1) into an alternative expression vector, pX459 (Ran, Hsu et al. 2013), which contains a puromycin resistance cassette. Using pX459 one can isolate successfully transfected cells with puromycin treatment, allowing determination of the rates of editing in a transfected population.

I therefore performed a series of transfections into non-immortalised heterozygous *mTmG* MEFs (as the immortalised MEFs are already puromycin resistant) using the pX459 LoxP plasmids and H2Bv2 with the Neon 100µl tips:

Well	Cell Number	LoxP#1 pX459 (ng)	LoxP#2 pX459 (ng)	H2Bv2 (ng)
1	$5x10^5$	0	0	0
2	$5x10^5$	5000	0	5000
3	$5x10^5$	0	5000	5000
4	$5x10^5$	2500	2500	5000

Cells were imaged on a Nikon A1R confocal microscope 24 hours post-transfection, at which point GFP-positive cells could be seen, indicating the CRISPR plasmids had begun to be expressed (Figure 3.23A). Therefore, the growth media was changed to MEF growth media + 3µg/ml puromycin. Forty-eight hours later the control cells had died, and so the media was replaced with antibiotic-free media in the remaining wells to allow cells to recover. Six days later, nine days post-transfection, wells were imaged in ten random positions at 20x on the Nikon A1R confocal. Images were quantified using the Cell Counter and Nucleus Counter plugins in Fiji to estimate rates of NHEJ. Unfortunately, as the cells were not immortalised, the primary MEFs doubled a limited number of times following transfection and the signal from the H2Bv2 plasmid was not diluted out, preventing accurate quantification of HDR. The results of this experiment indicated that approximately 70% of the transfected cells are visibly edited, with results somewhat higher for LoxP#1 than LoxP#2 pX459 (Figure 3.23C).


A
Pre-puromycin selection
No DNA
LoxP#1 pX459 + H2Bv2
LoxP#2 pX459 + H2Bv2
LoxP#1 + LoxP#2 pX459 + H2Bv2
B
Post-puromycin selection
LoxP#1 pX459 + H2Bv2
LoxP#1 pX459 + H2Bv2
LoxP#2 pX459 + H2Bv2
LoxP#1 + LoxP#2 pX459 + H2Bv2

C

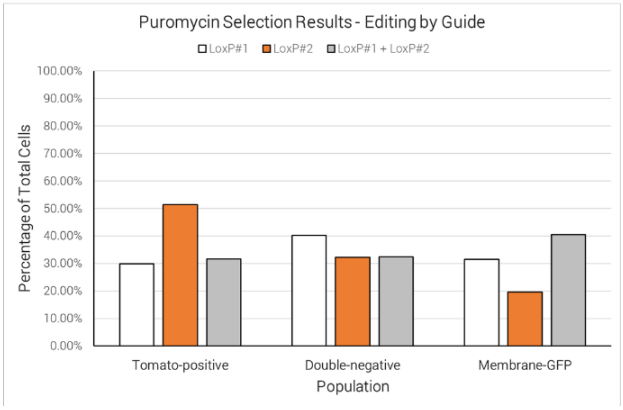

Puromycin Selection Results - Editing by Guide
LoxP#1
LoxP#2
LoxP#1 + LoxP#2
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Percentage of Total Cells
Tomato-positive
Double-negative
Membrane-GFP
Population

Figure 3.23 (on previous page): Following puromycin selection, the majority of transfected *mTmG* MEFs are visibly edited. A) Confocal images taken at 20x prior to puromycin selection. Scale bar 200µm. B) Confocal images taken at 20x post-puromycin selection and recovery. Scale bar 200µm. C) Ten 20x images were taken of each well in random positions and then quantified using the Nucleus Counter and Cell Counter plugins in Fiji.

3.16 Using Alternative Delivery Vectors to Improve Rates of Editing and Repair

The Neon system, with a transient plasmid delivery approach, has given highly variable results with regards to rates of editing between experiments, from 5% to approaching 40%. In order to improve reproducibility and minimize variability, I tested three additional methods for delivering the CRISPR components and repair templates:

a) **viral transduction**, taking advantage of the ability of lentivirus to deliver DNA into cells and adeno-associated virus (AAV) to stimulate repair;
b) **RNP transfection,** where ribonucleoprotein complexes consisting of ribonucleic acid and Cas9 protein are directly delivered, avoiding the need for transcription and translation; and
c) **minicircle transfection,** where the bacterial backbone has been removed (from the repair plasmids), allowing for a greater payload of repair template to be delivered and circumventing the cell's ability to recognise the repair template as foreign DNA and degrade it.

3.16.1 Viral Transduction Provides Mixed Results

A viral transduction was performed into heterozygous *mTmG* MEFs, using lentiviral delivery vectors (pseudotype: VSV-G) for Cas9 and guides, and an AAV vector to deliver the repair template (all received from Dr Peter Tennant, Section 2.12.4). I carried out eight separate transductions, using varying amounts of lentivirus (Lv) and AAV repair template:

Well	Cells Seeded	LoxP#1 Lv (µl)	LoxP#2 Lv (µl)	H2Bv2 AAV (µl)	Polybrene (µl)
1	$1x10^5$	0	0	0	1
2	$1x10^5$	0	0	4	1
3	$1x10^5$	4	0	0	1
4	$1x10^5$	0	4	0	1
5	$1x10^5$	2	2	0	1
6	$1x10^5$	4	0	4	1
7	$1x10^5$	0	4	4	1
8	$1x10^5$	2	2	4	1

The MOI (multiplicity of infection) for each virus preparation was:

LoxP#1 Lv – 362,
LoxP#2 Lv – 355, and
H2Bv2 AAV – 3760.

Following transduction, on day six, each well was imaged ten times at 20x magnification in random positions on a Dragonfly spinning disk confocal (Andor). Unfortunately, due to a software error, the images for control wells 1 and 2 were not saved, although no GFP-positive cells were observed in either well. Furthermore, few GFP-positive cells were seen in any of the LoxP Lv transduced wells (Figure 3.24), indicating a low rate of transduction. Further experiments using a viral transduction approach, performed by Dr. Peter Tennant, have since revealed highly variable rates of editing between different batches of lentivirus. For this reason, I did not pursue viral editing further.

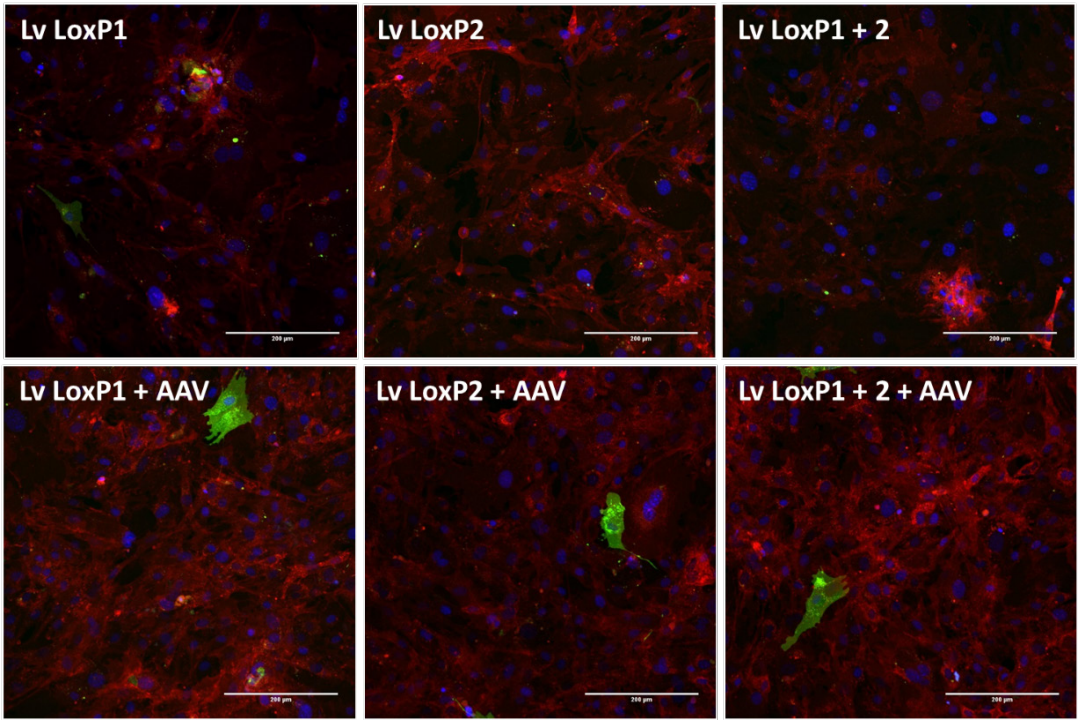


Figure 3.24: Lentiviral transduction of the heterozygous *mTmG* MEFs appears highly inefficient. Confocal images captured following transduction at 20x magnification. Scale bar 200µm.

3.16.2 High Editing Efficiencies can be Attained Using an RNP Approach

I prepared RNP complexes for both guides (Section 2.2.3), and then performed transfections using the Neon system and the 10µl electroporation tips into heterozygous *mTmG* MEFs (Section 2.12.1.3):

Well	Cell Number	LoxP#1 RNP (ng)	LoxP#2 RNP (ng)	Minicircle (ng)	Cas9 (µg)
1	$0.5x10^5$	0	0	0	1
2	$0.5x10^5$	240	0	0	1
3	$0.5x10^5$	0	240	0	1

Six days post-transfection, each well was imaged on a Dragonfly spinning disk confocal. Twenty-four hours later, the cells in each well were dissociated and prepared for FC, which was performed on a BD FACSJazz. A high editing efficiency could be seen with LoxP#1 RNP (~27%), however editing rates were much lower with LoxP#2 RNP (Figure 3.25); it is unclear why this would be the case, given the guides are the same between the plasmid, viral and RNP strategies. The FC plots again surprisingly

indicated that the largest proportion of the heterozygous (but immortalised) edited cells were double positive (Figure 3.25C).

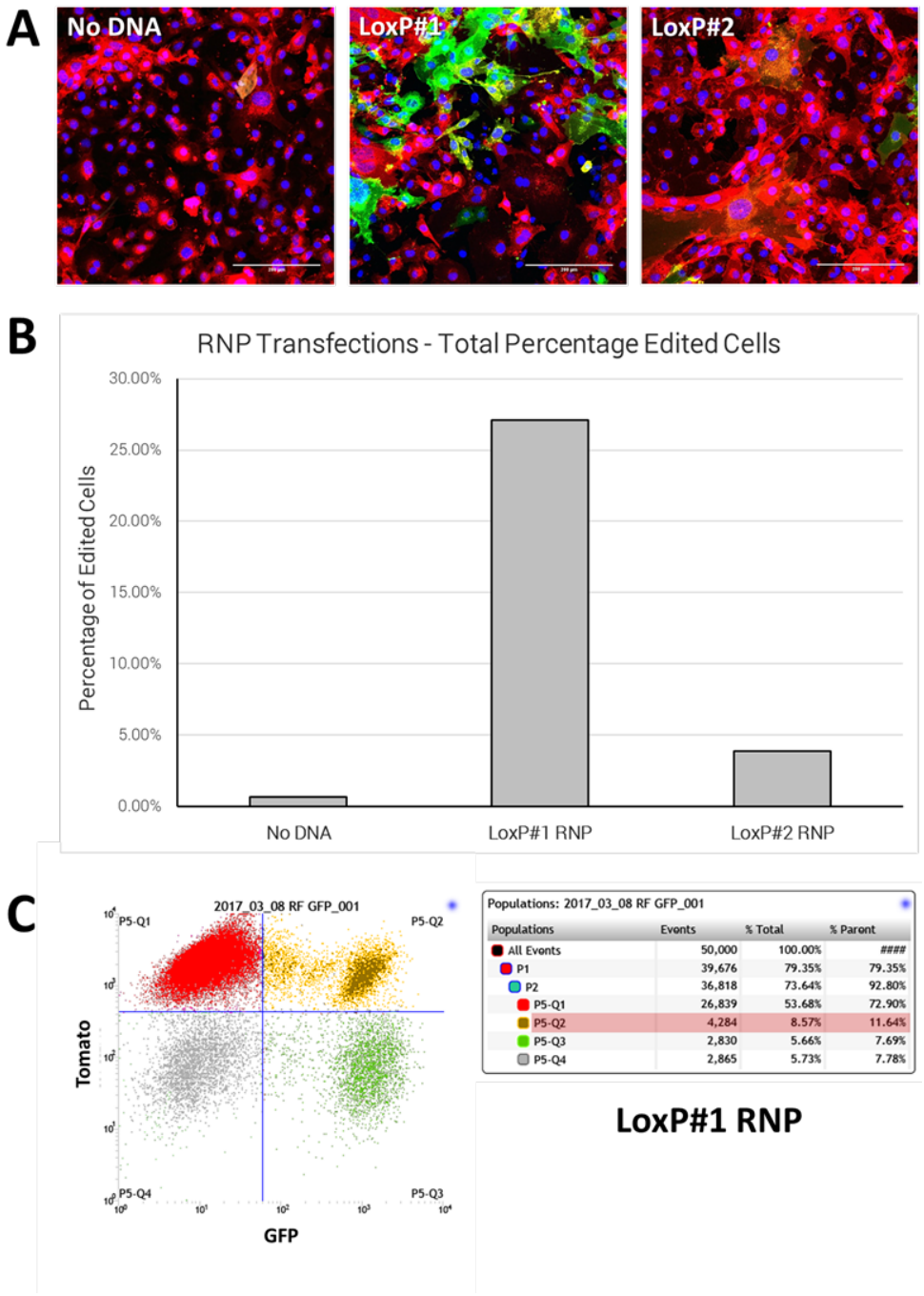


Figure 3.25: RNPs drive high rates of editing in immortalised heterozygous *mTmG* MEFs. A) Confocal images taken at 20x magnification, six days post-transfection. Scale bar 200µm. B) Each well was sorted seven days post-transfection on a BD FASCJazz and the total of the non-Tomato populations was plotted graphically. FC plots included on attached CD in file 'Figure 3_25'. C) FC plot for LoxP#1 RNP (seven days post-transfection) which shows the large double-positive population, at 11.64% (Q2).

3.17 The Double-positive Population in Immortalised *mTmG* MEFs May Be an Artefact of Genome Instability

Given the surprisingly large numbers of double-positive cells following an RNP transfection into heterozygous immortalised MEFs, it seems unlikely this population exists due to the half-life of the Tomato protein. A more plausible explanation is that several copies of the reporter emerge over time in our immortalised MEFs. We hypothesised that the process of immortalisation, in which we used SV40 large T-antigen to inactivate the tumour suppressor protein p53, could lead to much greater genome instability and an abnormal karyotype, with the potential outcome that our reporter locus is lost or duplicated. Less frequently, this may also be observed in late passage (>P8) primary MEFs which bypass senescence, by gaining spontaneous inactivating mutations in p53.

To test this hypothesis, I repeated the LoxP#1 RNP transfection in non-immortalised heterozygous *mTmG* MEFs at a low passage number (P3). In this iteration the minicircle repair template was also included. To ensure there were a large number of cells for the FC analysis, six transfections were performed (with the 10µl tips) for a total of $3x10^5$ cells, and the cells were returned to a 6-well glass-bottomed plate:

Well	Cell Number	LoxP#1 RNP (ng)	LoxP#2 RNP (ng)	Minicircle (ng)	Cas9 (µg)
1	$3x10^5$	0	0	0	6
2	$3x10^5$	0	0	3000	6
3	$3x10^5$	1,440	0	3000	6

Four days post-transfection, each well was imaged in 10 random positions at 20x using the Dragonfly spinning disk confocal. Imaging again revealed high rates of editing with the LoxP#1 RNP complex, including several cells with GFP-positive nuclei. Critically, there were no cells visible with GFP-positive nuclei in the minicircle only control well (Figure 3.26A). Four days after imaging, wells were analysed using a BD FACSAria. FC results revealed, as expected, a very small proportion, 0.3%, of double-positive cells following this transfection (Figure 3.26B); indicating immortalisation leads to genome instability affecting our reporter locus. Furthermore, editing rates were again very high (approaching 40%) and several cells with GFP-positive nuclei

could be seen (arrows, Figure 3.26A, C), suggesting that this approach may be more efficient and reliable than plasmid delivery. Despite the potential genome instability, given the ease of handling and scalability of the immortalised MEFs, I have continued to utilise them for specific applications.

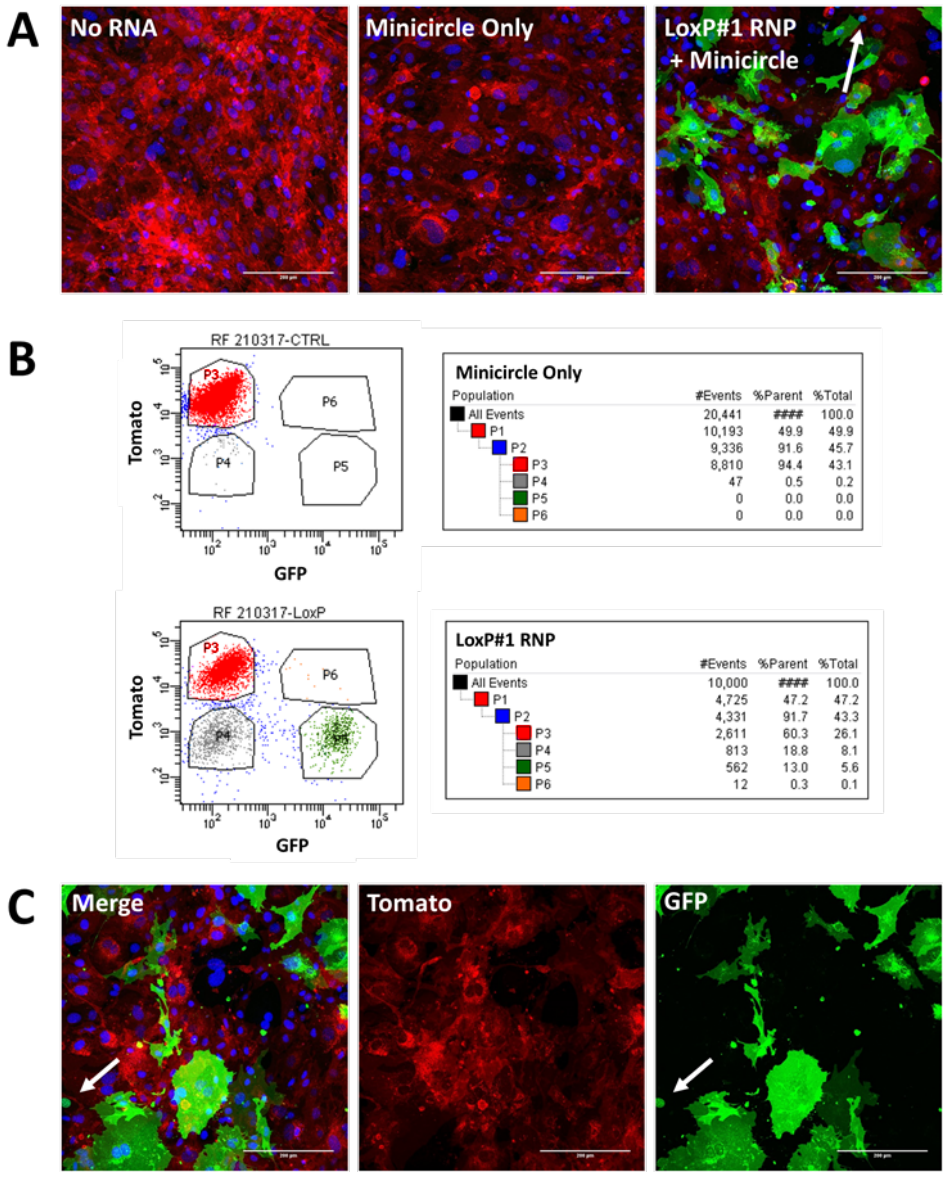


Figure 3.26: Double-positive cells are not detected in non-immortalised MEFs following RNP transfection. A) Cells were imaged at 20x on a confocal microscope four days post-transfection. Scale bar 200µm. B) Eight days post-transfection each well was analysed on a BD FACSJazz. Plots for the minicircle control and the LoxP#1 RNP + minicircle wells are shown. C) Viewing the LoxP#1 + minicircle confocal images with split channels reveals the lack of Tomato fluorescence in several of the GFP-positive cells, although this is more evident on the FC plots due to the high confluency of the wells. Scale bar 200µm. White arrows indicate cells with GFP-positive nuclei.

3.18 Determining the Kinetics of Editing in the *mTmG* System Using Time-lapse Imaging

In order to determine how quickly our system responds to editing events, via the production of EGFP and degradation of Tomato, I carried out a time lapse imaging experiment following transfection. In addition, time-lapse imaging will allow us to confirm that both proteins are present for the entirety of the cell cycle, especially during mitosis/cytokinesis. This will reveal whether the existence of a transient negative population could explain the mis-sorting we see by FACS.

3.18.1A 48-hour Time-Lapse Post-Transfection Reveals the Speed of the *mTmG* System

Using the Dragonfly confocal microscope, a time-lapse was set up, this time for 48 hours. Four wells of a 6-well plate were transfected using the Neon 100µl tips (Section 2.12.1.1):

Well	Cell Number	Cas9 Plasmid(s)	Cas9 Plasmid DNA (ng)	Repair	Repair DNA (ng)
1	$2x10^5$	N/A	0	N/A	0
2	$2x10^5$	LoxP#1 pX330, LoxP#2 pX330	2000	N/A	0
3	$2x10^5$	LoxP#1 pX330, LoxP#2 pX330	2000	Minicircle	2000
4	$2x10^5$	LoxP#1 pX330, LoxP#2 pX330	2000	Minicircle	2000

$2x10^5$ heterozygous *mTmG* immortalised MEFs were transfected per well to ensure the cells were reasonably sparse for the duration of the time-lapse. Cells were imaged 6 hours post-transfection. A multi-position programme was used to image each well in 10 positions, taking a 10µm Z-stack in all channels every 20 minutes. A maximum intensity projection was then taken, and the images were compiled to form videos (40 in total).

The resulting time-lapse videos revealed that editing is rapid in our system, even when using plasmids which would need to be transcribed and translated. Using these videos, I tracked three cells which began to show EGFP expression within ~24 hours post-transfection and completed the transition from Tomato to EGFP within 48 hours

(Figure 3.27, 3.30, NB. similar kinetics can be seen in the videos in many other MEFs however it is only possible to cleanly perform this analysis on the very small proportion of MEFs which remain separate from all other cells for the entire time-lapse due to overlapping fluorescence). There appears to be a peak in editing, based on the number of cells beginning to express GFP, between 30 and 36 hours post-transfection. Unfortunately, as imaging positions were randomised, very few cells were captured which expressed nuclear-GFP thus it is difficult to comment on the kinetics of repair, although one would expect it to occur across a similar time-frame. Several of the resulting videos are included in the attached media CD (saved as Time-course XX.avi, referred to in the following sections).

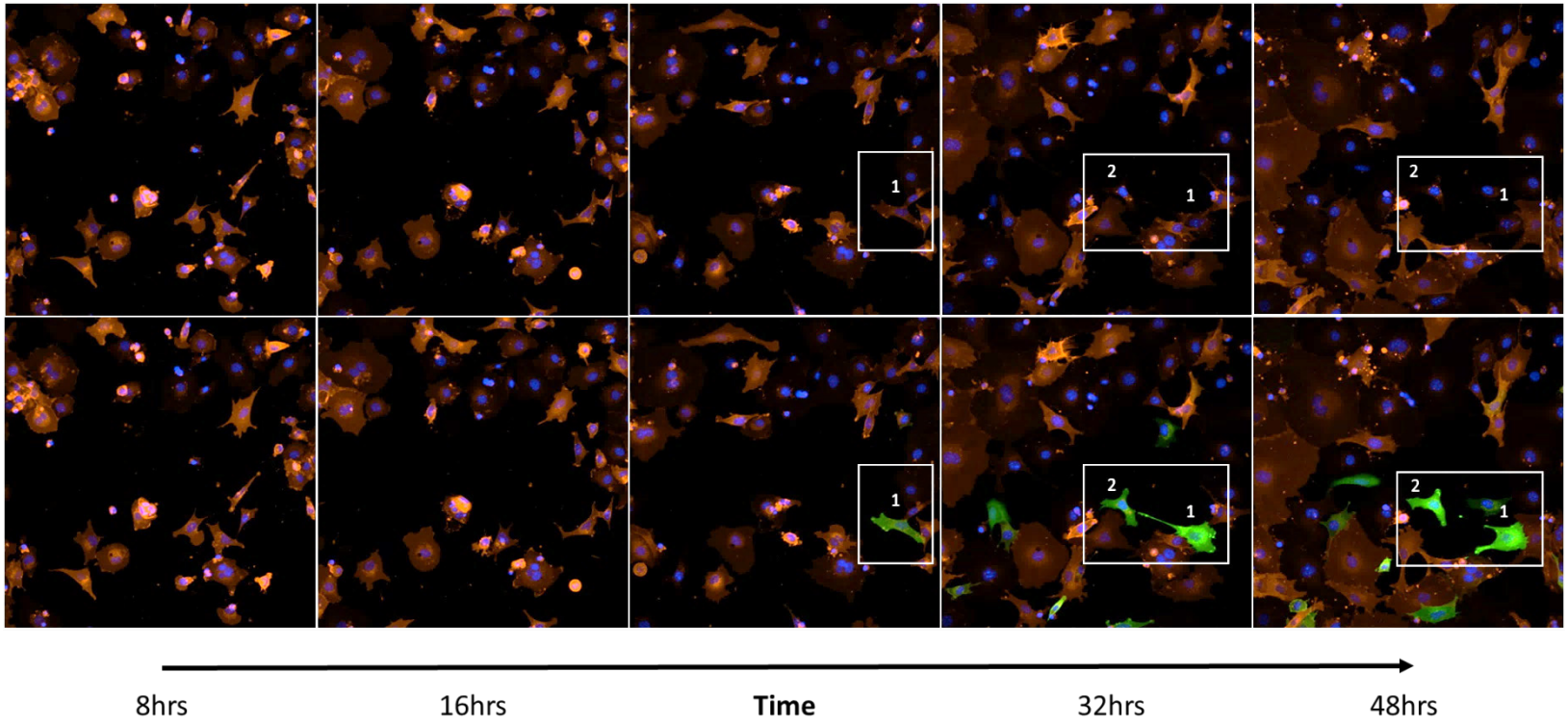


Figure 3.27: Time-lapse imaging in MEFs reveals the kinetics of EGFP production following editing and Tomato degradation, often reaching completion within 48 hours post plasmid transfection. Images were captured at 20x magnification every 20 minutes, beginning 6 hours post-transfection and continuing for 48 hours. Above: Tomato and Hoechst, below: merge. Two cells are highlighted which gain membrane-GFP and lose membrane-Tomato. The video this panel is derived from is included in the attached media CD (Time-course 08.avi).

3.18.2 Quantification of Kinetics Using Imaris

In order to quantify the time-lapses captured, and to gain objective measurements of the change in fluorescence intensity, the Track (cells) module in Imaris was used, with the settings detailed in Section 2.16.2.

3.18.2.1 Quantifying Background Fluorescence

Firstly, I quantified background fluorescence by identifying regions with cells in the image, and then averaging the fluorescence intensity detected outside these regions. This analysis revealed a remarkably constant background fluorescence detected throughout the whole time-lapse, with a reading in both EGFP and Tomato channels of ~120 arbitrary units (AU) (Figure 3.28).

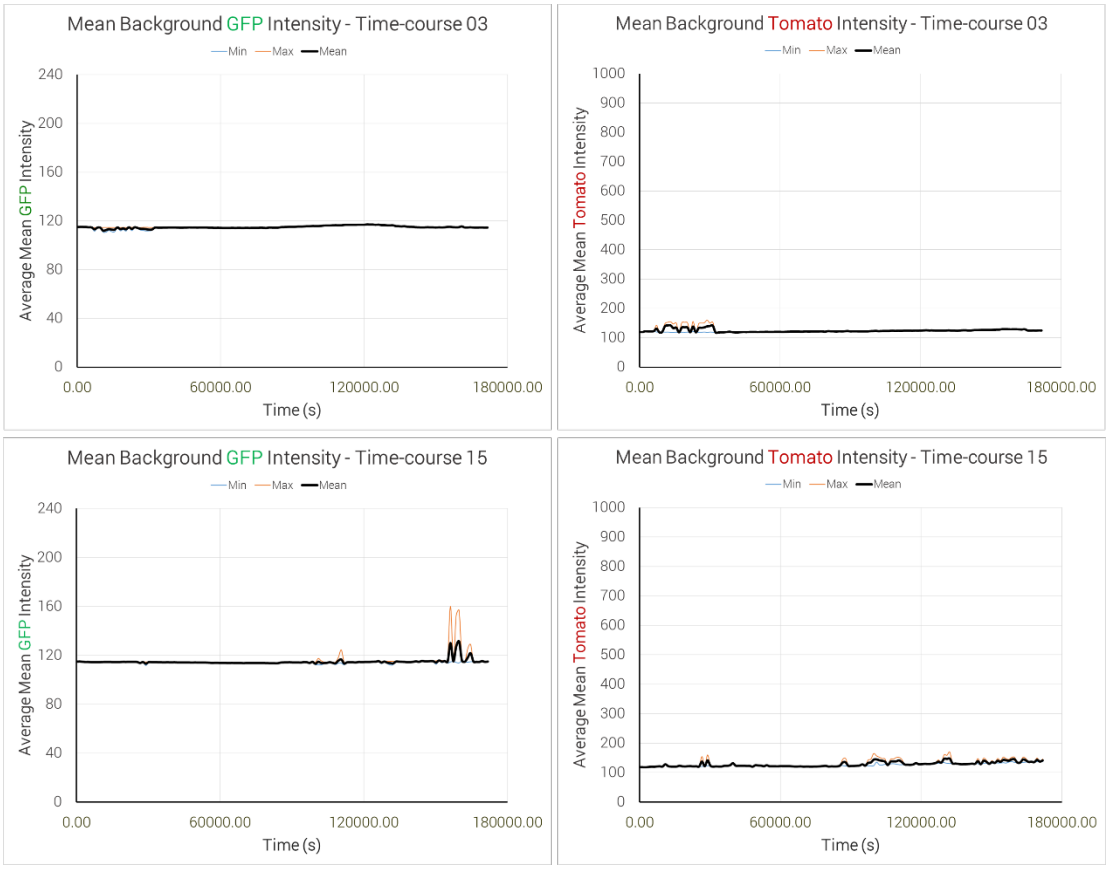


Figure 3.28: Cell-free background fluorescence intensities in the GFP and Tomato channels are constant. The mean fluorescence intensities in regions outside cells was calculated at every time point during two time-lapse videos using Imaris. Axes were chosen for comparison with Figure 3.29. The videos for Time-course 03 (above) and Time-course 15 (below) are included on the attached media CD.

3.18.2.2 GFP Fluorescence Does Not Rise Above Background Levels in Non-Transfected Cells

As a second control, I analysed time-lapses from the non-transfected, electroporated sample (well 1) to gauge the variation in EGFP and Tomato fluorescence intensities during imaging. Imaris Track was again used to separate the videos into individual cells, and then the mean fluorescence intensity across the volume of each cell was measured. The variation in the mean fluorescence intensities was then plotted in Figure 3.29. EGFP fluorescence could be seen to be constant and equal to the background fluoresce seen in the previous section; that is ~120 AU. Tomato fluorescence was more variable and averaged between 300 and 400 AU, with the mean intensity for any single cell not dropping below 200 AU.

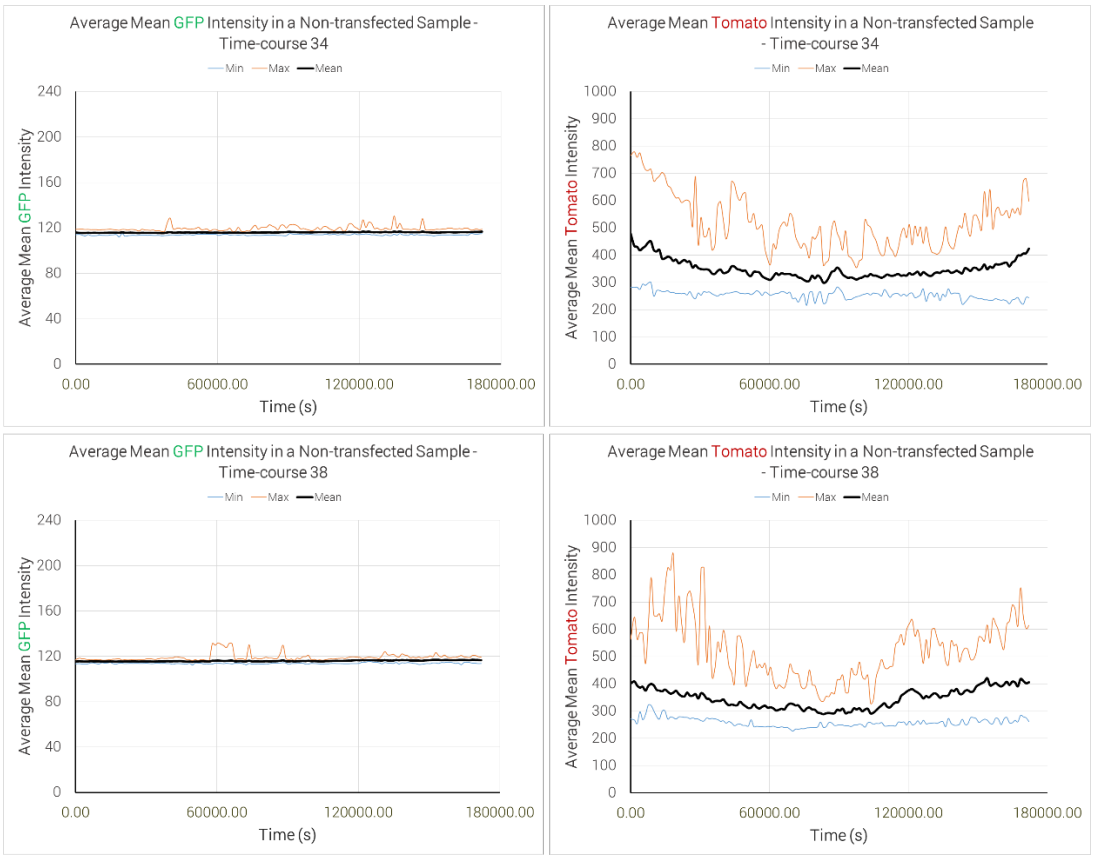


Figure 3.29: In non-transfected *mTmg* MEFs, GFP fluorescence remains at background levels with constant Tomato fluorescence. The mean fluorescence intensity across the volume of each cell was calculated by Imaris. The minimum, maximum and mean values for the population of detected cells was then plotted graphically. The videos for Time-course 34 (above) and Time-course 38 (below) are included on the attached media CD.

3.18.2.3 Transfected Cells Display a Rapid Switch in Fluorescence from Tomato to EGFP

In order to measure the change in fluorescence in transfected cells, Imaris Track was again used, this time to identify cells based on their GFP fluorescence intensities. Once a cell was detected as being GFP-positive, the mean GFP and Tomato intensities across the volume of the cell were tracked for the duration of the time-course. The results revealed that edited cells display a rapid decrease in Tomato intensity across a period of 24 hours, from approximately 500 to 200 AU or below, while there is a simultaneous substantial increase in GFP intensities (Figure 3.30).

Across the 40 videos there were many cells which appear to be double-positive following transfection (not quantified), in the time-frame of 54 hours post-transfection. Interestingly, a double-positive cell from Time-course 08 can be tracked as it undergoes a cell division following developing GFP fluorescence (Figure 3.31A). By tracking the mean Tomato and GFP intensities of this cell as it undergoes cell division, one can see that there is no period during the cell cycle in which fluorescence from either reporter is lost (Figure 3.31B). Other instances of Tomato- and GFP-positive cells dividing in the included time-courses confirm this finding; which indicates that the double-negative population seen following editing cannot be a cell cycle artefact. Furthermore, by tracking the sum intensities in both channels, one can clearly see when the cell is undergoing cytokinesis, and then when the two daughter cells are produced (Figure 3.31C).

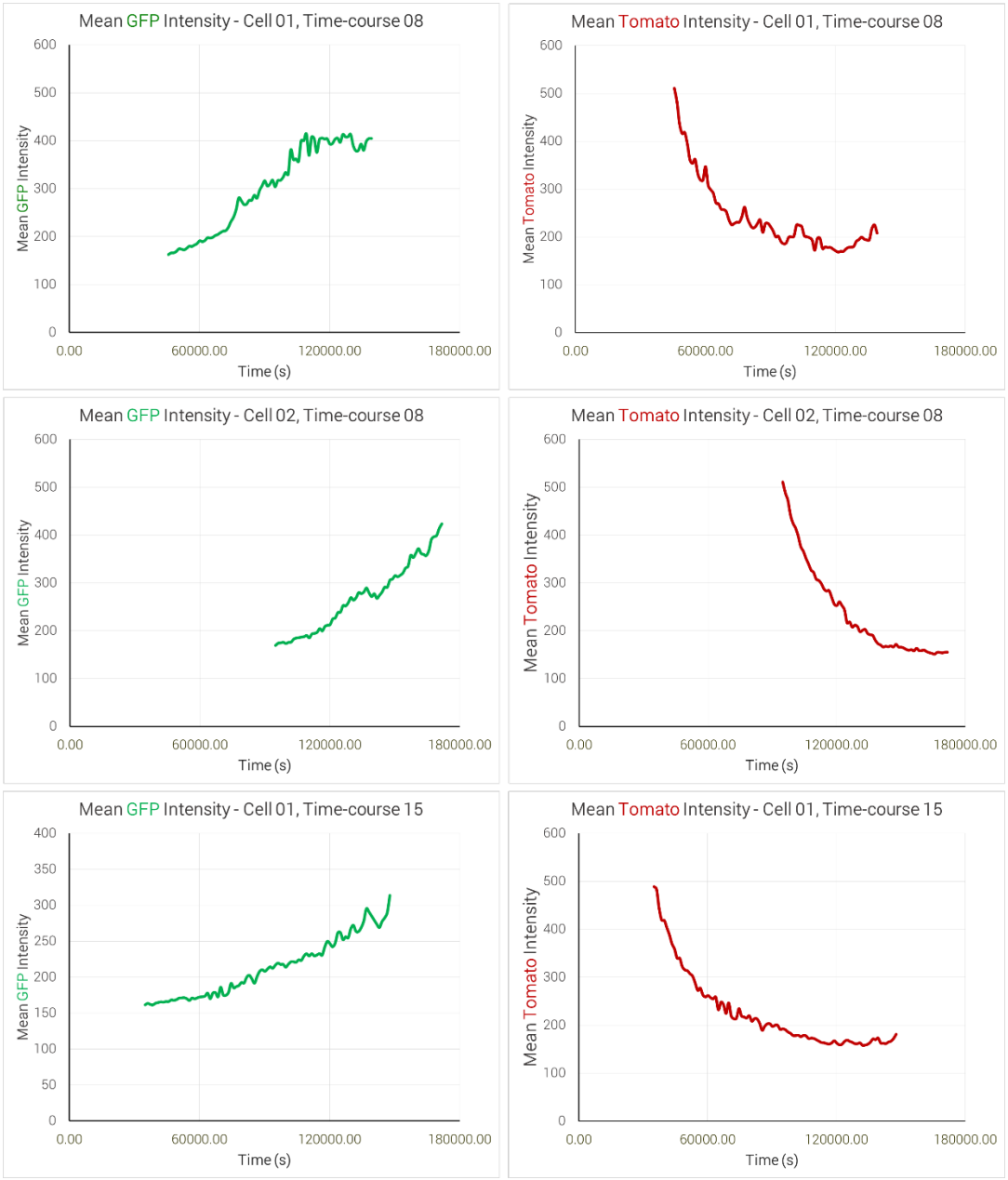


Figure 3.30: Edited mTmG MEFs display a rapid switch in fluorescence from Tomato to GFP within 24 hours. Imaris Track was used to identify cells based on GFP fluorescence, and then the mean GFP and Tomato intensity across the volume of the cells was calculated for each time-point. Here, Cell 01 and Cell 02 from Time-course 08 correspond to the cells highlighted in Figure 3.27. Videos for Time-course 08 and Time-course 15 are included on the attached media CD.

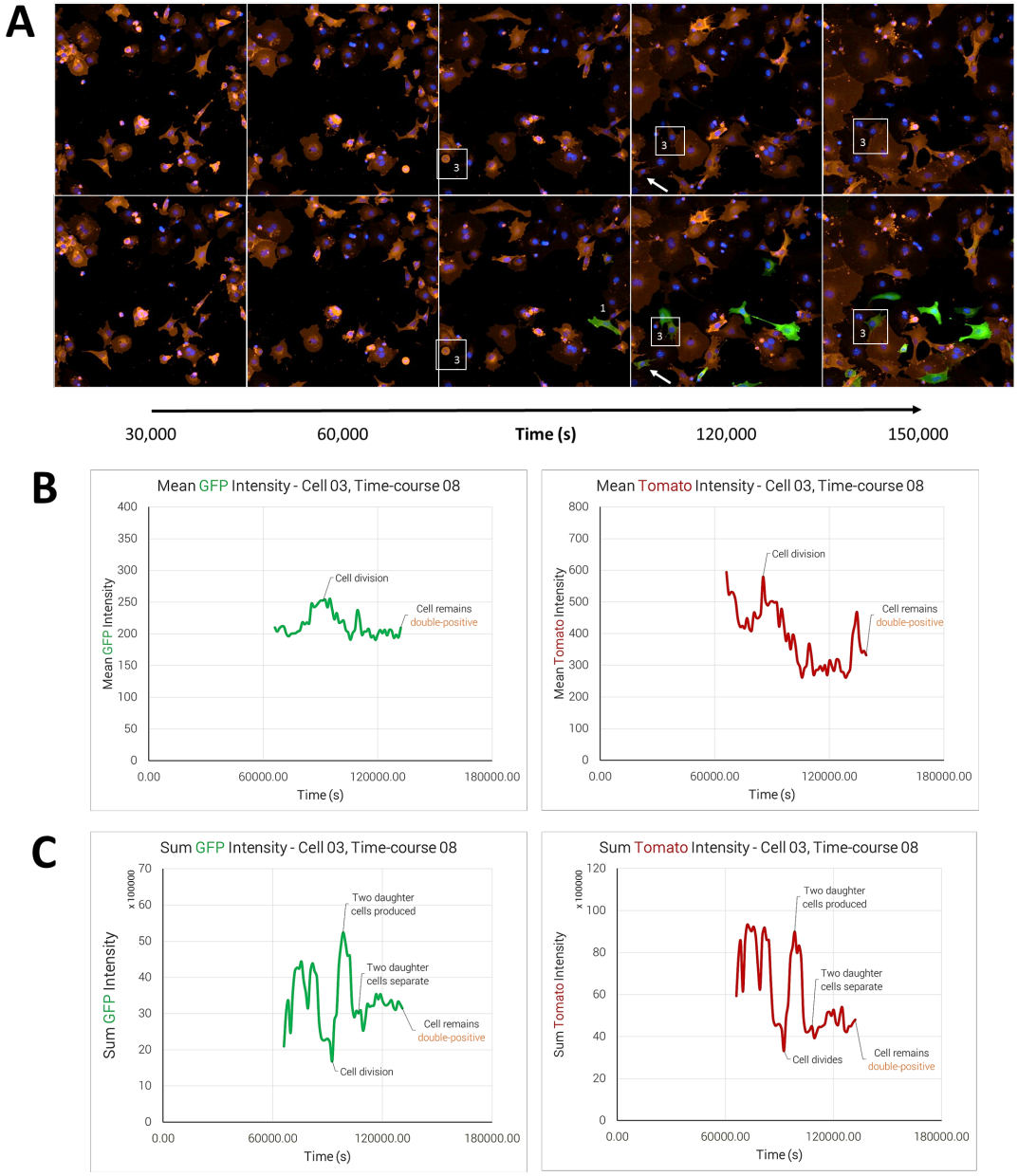


Figure 3.31: Both Tomato and GFP fluorescence are maintained throughout the cell cycle in *mTmG* MEFs. A) One double-positive cell, Cell 03, was tracked for the duration of Time-course 08, as it undergoes cell division soon after gaining GFP fluorescence. The daughter cell produced is indicated with an arrow. B) Plotting the mean intensities of the cell as it undergoes cell division reveals that there is no phase of the cell cycle in which fluorescence is lost or drastically reduced. C) Plotting the sum intensities of the detected object clearly highlights several points during cell division. The video for Time-course 08 is included on the attached media CD.

3.19 The Small Molecules NU-7441 and Mirin Provide Unexpected Results in the *mTmG* System

Previous pilot trials with small molecules to improve rates of HDR have proven inconclusive (Section 3.14). This may be due to two reasons; our limited ability to capture a large, representative sample by imaging to get accurate counts; and/or due to inherent differences in the repair machinery employed between the cell types used in their respective published studies and in our MEF model. During the course of my PhD it became possible to partially address both of these questions.

In terms of imaging, a novel platform became available, the Zeiss Celldiscoverer 7, which allows the quick, automated capture of the majority of a well in several channels. Following imaging, the associated Zen software is then able to detect and quantify various parameters, including the percentage of cells with GFP-positive nuclei. This platform allowed us to sample a vast number of cells, approximately 140,000 per well in a short space of time, potentially providing a more reliable and consistent figure for HDR. This platform was used to test a recently described small molecule, NU-7441, which had been shown to specifically increase rates of HDR in MEFs (Robert, Barbeau et al. 2015). I carried out three separate transfections with the Neon 100µl tips (Section 2.12.1.1) for comparison:

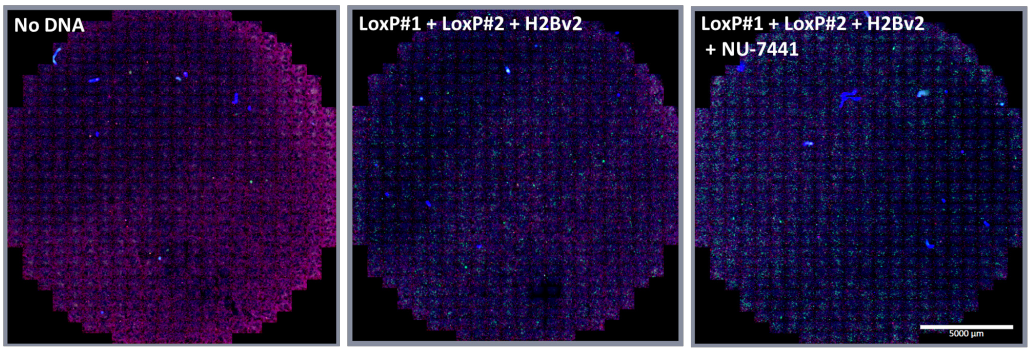
Well	Cell Number	Cas9 Plasmid(s)	Cas9 Plasmid DNA (ng)	Repair Plasmid	Repair DNA (ng)	Small Molecule
1	$5x10^5$	N/A	0	N/A	0	N/A
2	$5x10^5$	LoxP#1 pX330, LoxP#2 pX330	5000	H2Bv2	5000	N/A
3	$5x10^5$	LoxP#1 pX330, LoxP#2 pX330	5000	H2Bv2	5000	NU-7441 (2µM)

An optimised concentration of 2µM NU7741 was used following the protocol by Robert, Barbeau et al. (2015), where the small molecule was added immediately following transfection and maintained for 24 hours. Transfections were performed using the Neon 100µl tips (section 2.12.1.1), into immortalised heterozygous *mTmG* MEFs. Five days post-transfection, all three wells were split 1 in 2 in order to disrupt any clonal outgrowth. The following day, a 10x scan was taken of the central 25% of each well of the 6-well plate (Figure 3.32A). A bespoke automated analysis pipeline

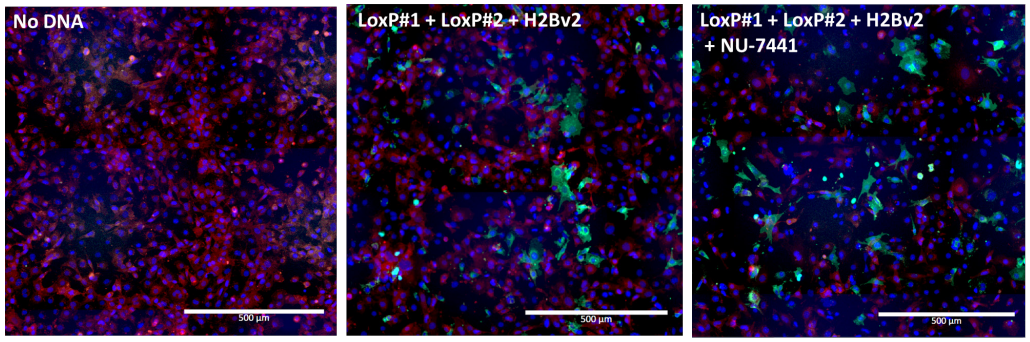
was constructed in the Zen software which first applied a mask based on DAPI intensity (with a watershed) to provide a total cell count, and then a second mask based on GFP intensity, area and circularity to provide a count of the number of GFP-positive nuclei (Figure 3.32C). This allowed us to report the percentage of GFP-positive nuclei in each well. Following imaging, 24 hours later, each well was dissociated and analysed by FC to gain a read out of the total visible editing in each well (Figure 3.32D). This was used to normalise for any difference in transfection efficiencies between the two wells. The results indicated that there was potentially a greater than 2-fold increase in HDR when treating with NU-7441 (Figure 3.32E).

Unfortunately, the Celldiscoverer 7 was only available on-site for a brief demo period. Therefore I decided to attempt to repeat and confirm the results using the Nikon A1R confocal microscope, taking advantage of a machine learning tool, CellProfiler Analyst (Jones, Kang et al. 2008), to quickly score a large number of images for nuclear-GFP. As before, heterozygous *mTmG* immortalised MEFs were transfected with the LoxP pX330 plasmids and the H2Bv2 repair plasmid with or without 2μM NU-7441 for 24 hours, however this time three biological replicates were carried out for each condition (*mTmG* MEFs were split into three batches which were transfected, treated, maintained and assayed separately). Seventy-two hours post-transfection, each well was split 1 in 2. One half of the split was returned to the plate for imaging, and the other half was taken for analysis on a BD LSRFortessa to determine editing rates. Twenty-four hours later, four days post-transfection, each well was imaged at 20x on the A1R in a 10x10 grid array (for the equivalent of 100 fields at 20x, Figure 3.33A, B). CellProfiler Analyst (Section 2.16.3) was then used to calculate the rates of HDR seen in each image. Unfortunately, using this methodology, no significant difference could be seen in HDR rates upon NU-7441 treatment, however, there was a significant increase in the number of edited cells seen by FC ($p < 0.05$, Figure 3.33C).

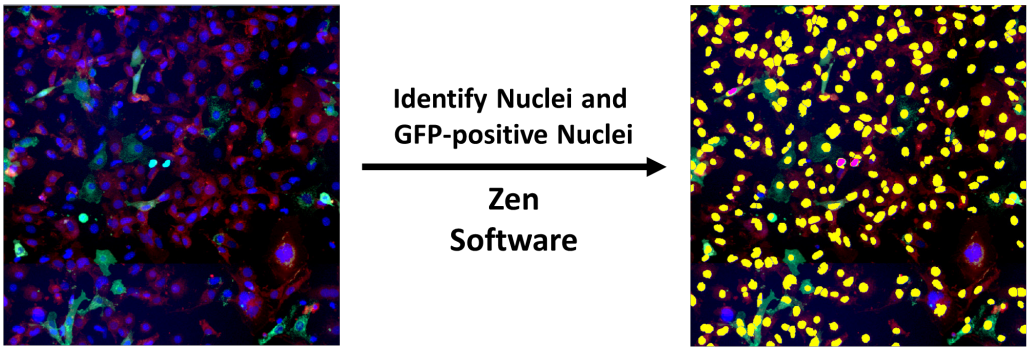
A



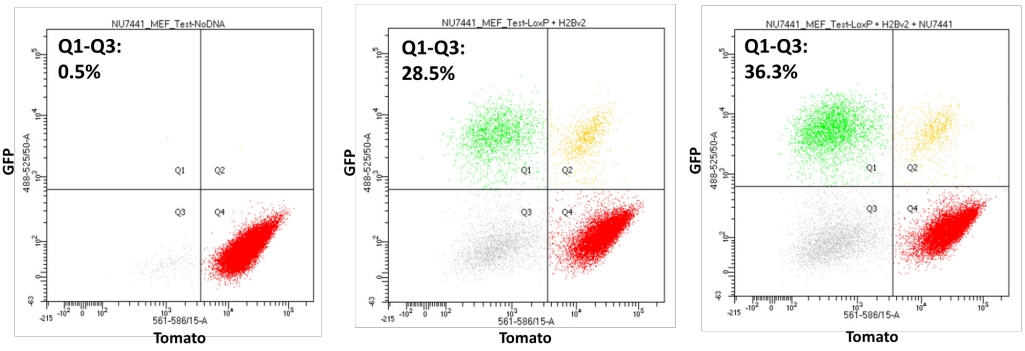
B



C



D



E

GFP-positive Nuclei in:	No DNA	LoxP + H2Bv2	LoxP + H2Bv2 + NU-7441
As a Percentage of Total Cells	0.00%	0.43%	1.14%
As a Percentage of Edited Cells	0.00%	1.51%	3.14%

Figure 3.32 (on previous page): The small molecule NU-7441 appears to increase HDR in *mTmG* MEFs following plasmid transfection. A) A 10x scan was taken of the central 25% of each well of a six well plate using a Zeiss Celldiscoverer 7, five days post-transfection. Scale bar 5000µm. B) Enlargements of the above images reveal a high level of detail, including nuclei and GFP-positive nuclei. Scale bar 500µm. C) The Zen software is able to process the images to automatically identify nuclei (in yellow) and GFP-positive nuclei (in purple). D) Each well was analysed by FC following imaging to estimate transfection efficiencies as a proxy of editing rates (noted in the corner of the plots). E) The results of the automated counting of nuclei and GFP-positive nuclei produced by the Zen software, presented as rates of HDR with or without normalisation to the editing efficiencies derived from FC.

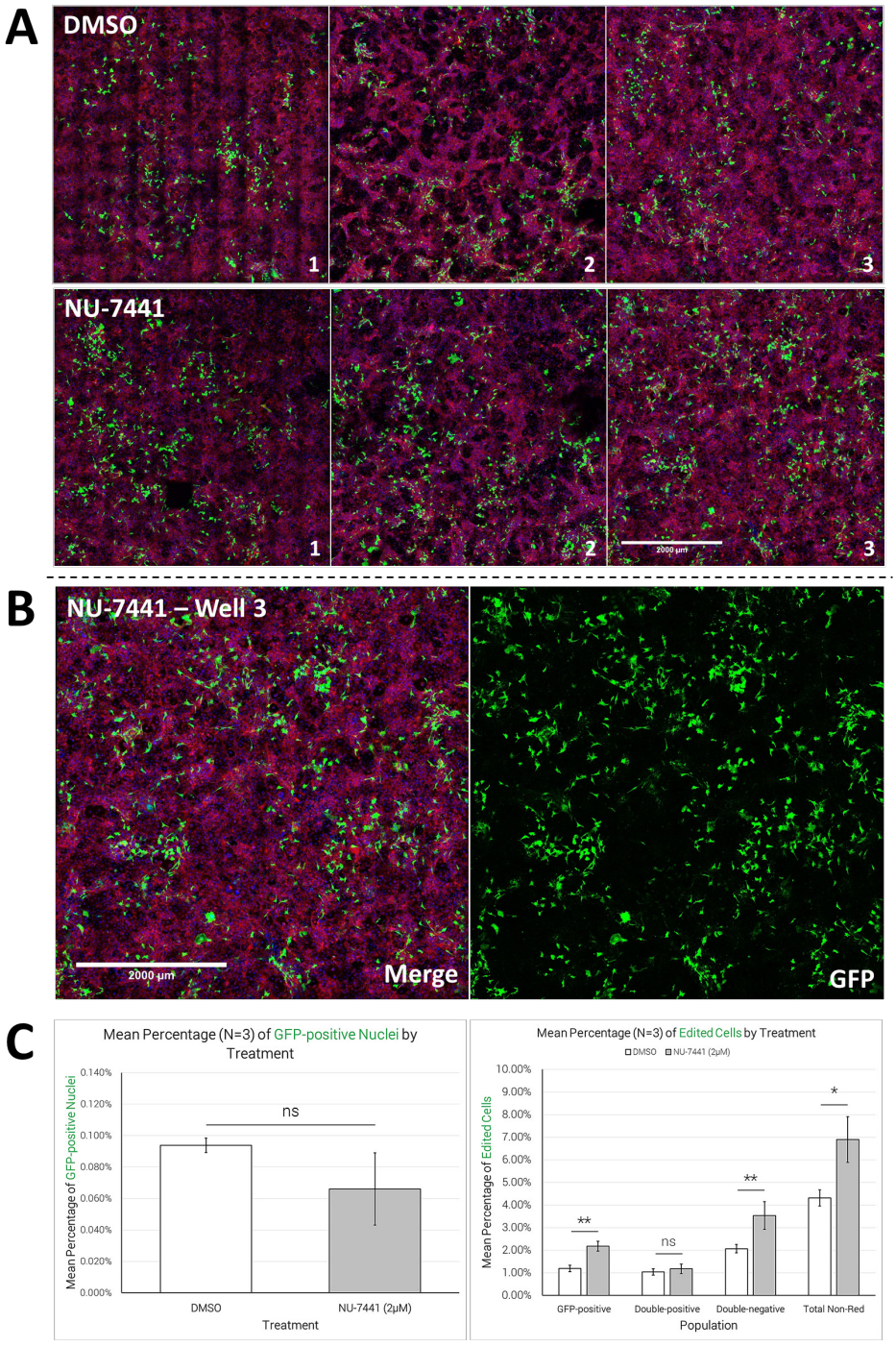


Figure 3.33: Using CellProfiler Analyst quantitation, the increase in HDR on NU-7441 treatment in *mTmG* MEFs failed to replicate, but FC did show a statistically significant increase in total editing. A) A 10x10 grid was imaged in each well at 20x using a Nikon A1R confocal microscope. Scale bar 2000µm. B) An enlargement of the image for NU-7441 well 3 displays a good level of detail, however significantly fewer cells have been captured than with the Celldiscoverer 7. Scale bar 2000µm. C) CellProfiler results (left) and FC results (right) with the corresponding statistical significances from a Student's t-test indicated (N=3 for each condition). FC plots included on attached CD in file 'Figure 3_33'.

3.19.1 The Increase in Editing upon NU-7441 Treatment is Maintained in the Absence of a Repair Template

In order to rule out the possibility that the increase in editing we are seeing with NU-7441 treatment is due to repair, I repeated the previous experiment with the omission of the H2Bv2 repair template. Three biological replicates were prepared for each condition, that being with or without NU-7441 treatment, and 72 hours post-transfection each well was dissociated and analysed on a BD LSRFortessa. The results again indicated a statistically significant increase in the percentage of edited cells, this time in all quadrants, following a 24-hour incubation with 2μM NU-7441 (Figure 3.34).

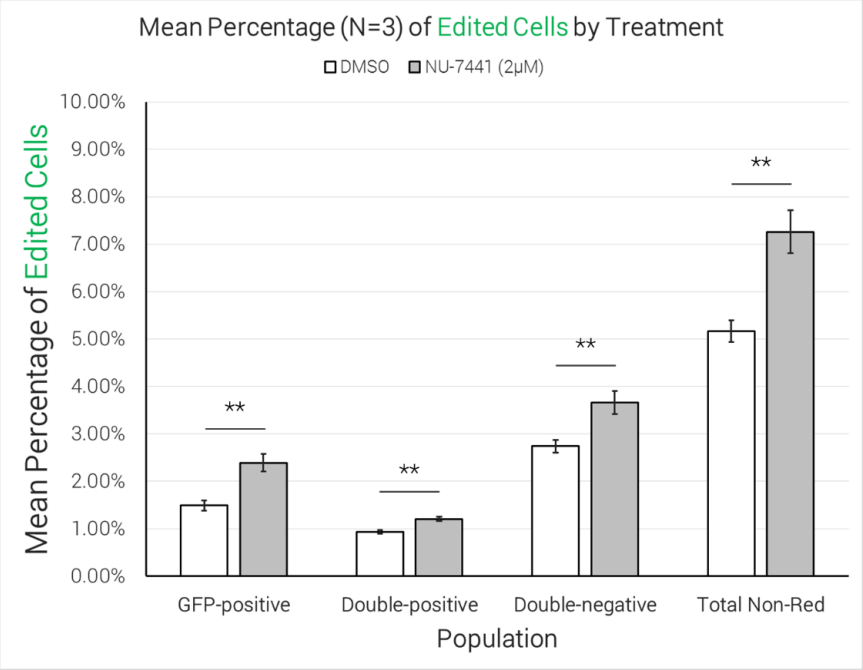


Figure 3.34: In the absence of a repair template, NU-7441 treatment is still able to drive a statistically significant increase in editing in *mTmG* MEFs. Error bars represent the standard deviation of three biological replicates and stars indicate the statistical significance from a Student's t-test (N=3). FC plots included on attached CD in file 'Figure 3_34'.

3.19.2 The Increase in Editing Does Not Appear to be Driven by MMEJ

As a potential explanation for the increase in editing seen with NU-7441 treatment, I noted that there would be small regions of homology remaining on either side of the loxP sites following cleavage with either guide (Figure 3.3). These regions may allow repair by the microhomology-mediated end joining pathway (MMEJ), as opposed to

NHEJ or HDR. Furthermore, it has been shown that blocking the canonical NHEJ pathway may promote the use of MMEJ (Sharma, Javadekar et al. 2015). To examine whether we are seeing increased rates of editing due to the promotion of the MMEJ pathway in our cells, I repeated the previous experiment with the addition of the small molecule Mirin, which inhibits MMEJ (and HDR) but has no effect on NHEJ (Sharma, Javadekar et al. 2015). Mirin was used at a concentration of 50µM as this concentration has previously been shown to result in 95% inhibition of MMEJ in HEK293 cells (Dupré, Boyer-Chatenet et al. 2008).

The experiment was repeated as before with the addition of Mirin at 50µM to the wells with NU-7441, and an additional three wells with Mirin (50µM) alone. Cells were again allowed to proliferate for 72 hours post-transfection prior to being analysed on a BD LSRFortessa. The FC results indicated that Mirin treatment alone did not appear to be having a significant effect on editing, however there was still a significant increase in editing on treating with both NU-7441 and Mirin (Figure 3.35).

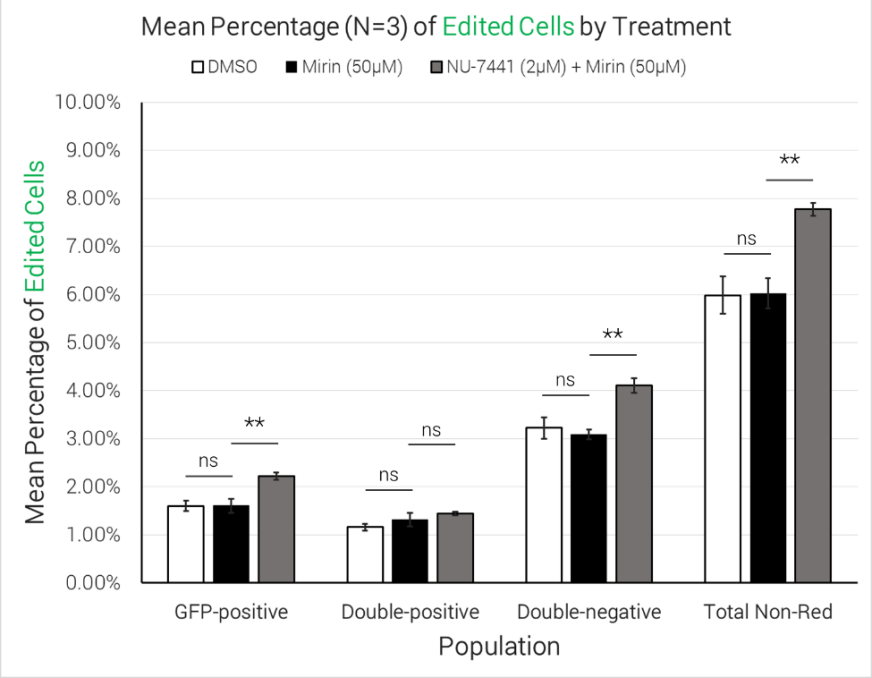


Figure 3.35: The increase in editing in *mTmG* MEFs seen upon NU-7441 treatment does not appear to be driven by MMEJ, as a potent MMEJ inhibitor, Mirin, fails to alter editing outcomes. Here, error bars represent the standard deviation of three biological replicates and stars indicate the statistical significance from a Student's t-test. Upper bars indicate the statistical significance between NU-7441 + Mirin and No Small Molecule. FC plots included on attached CD in file 'Figure 3_35'.

3.19.3 Using RNPs to Limit the Duration of Editing Does Not Affect the Increase Seen upon NU-7441 Treatment

As it is known that plasmid expression persists for several days (DeWitt, Corn et al. 2017), it is possible that we are missing the window of editing when treating with the small molecules for twenty-four hours. To correct for this, I decided to repeat the experiments with RNP complexes instead of plasmids, as these will immediately be active once inside the cells, and thenceforth will be rapidly degraded. Indeed, Cas9 RNPs have been shown to be almost entirely absent at 24 hours post-transfection, as opposed to plasmid-derived Cas9 which reaches its peak abundance at this time point (DeWitt, Corn et al. 2017).

RNP transfections were performed using the Neon into heterozygous *mTmG* immortalised MEFs (Section 2.12.1.3), using a 1:1 mixture of the LoxP#1 RNP complex and the LoxP#2 RNP complex:

Well	Cell Number	LoxP#1 RNP (ng)	LoxP#2 RNP (ng)	Minicircle (ng)	Cas9 (µg)
1 - 12	$0.5x10^5$	120	120	0	1

Following transfection, the cells were immediately transferred into growth media containing the small molecules as before and incubated for a total of 24 hours. Within this window, a large amount of cell death was seen in the wells containing the small molecules, so the cells were allowed to proliferate for a further five days prior to being analysed by FC. In order to gauge the effect the small molecules are having on cell viability, each well was re-suspended in the same volume of PBS and then sorted for exactly 120 seconds. The total number of cells sorted from each well was then recorded.

The FC results returned following RNP transfection were broadly similar to those following plasmid transfection (Figure 3.36). There was a significant increase in editing upon incubation with NU-7441 which was not reduced by the addition of Mirin. However, as opposed to the previous experiments, there appeared to be a small increase in editing when incubating with Mirin alone following the RNP transfections. Furthermore, both small molecules proved to be cytotoxic following a 24-hour

incubation at the concentrations employed and were very highly cytotoxic when combined (Figure 3.37).

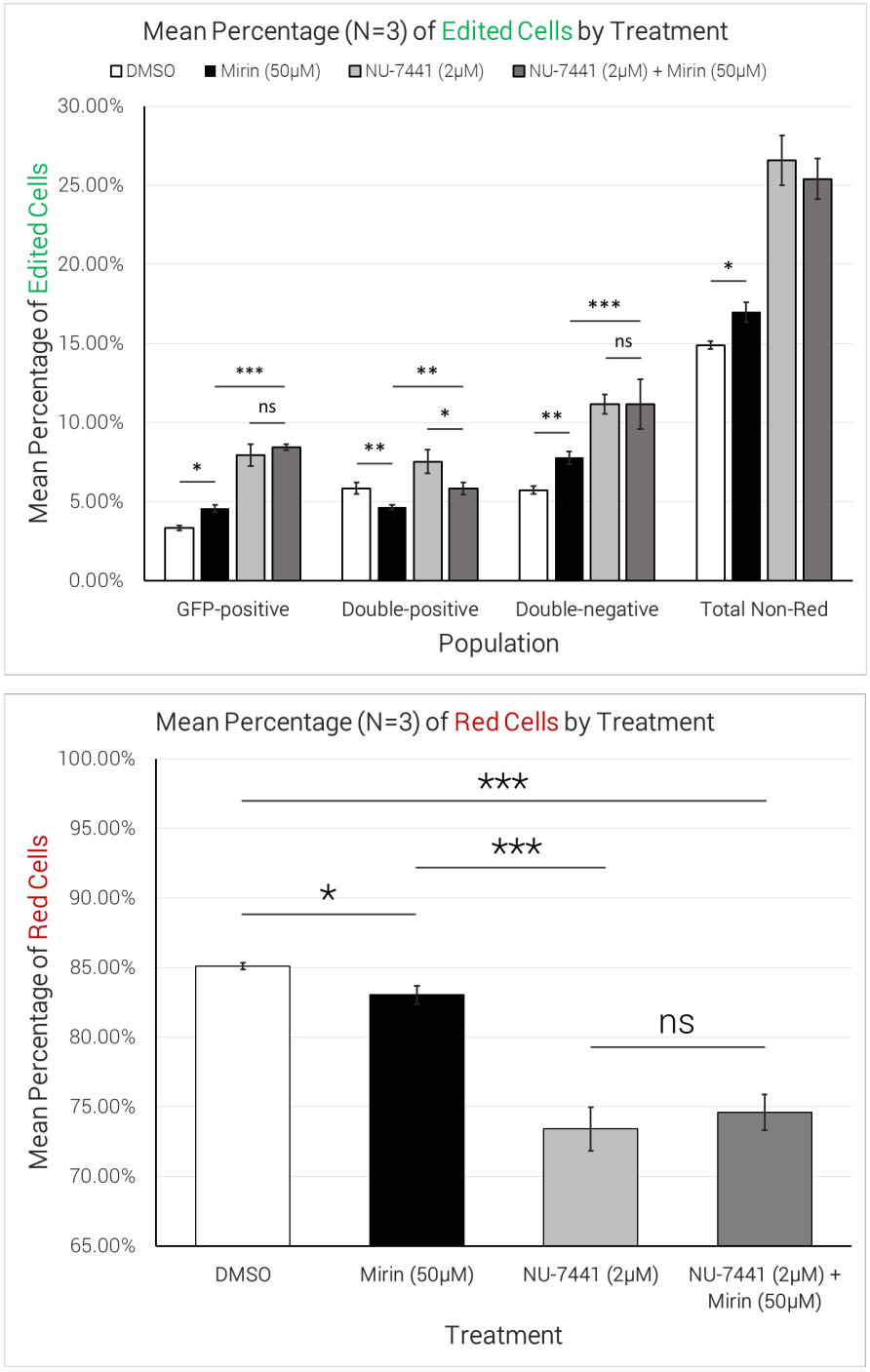


Figure 3.36: The increase in editing is also seen upon RNP transfection, following incubations with NU-7441 in *mTmG* MEFs. Error bars represent the standard deviation of three biological replicates and stars indicate the statistical significance from a Student's t-test. NB. Total Non-Red in the above graph is the inverse of the graph below (red cells), which has been shown for clarity. FC plots included on attached CD in file 'Figure 3_36'.

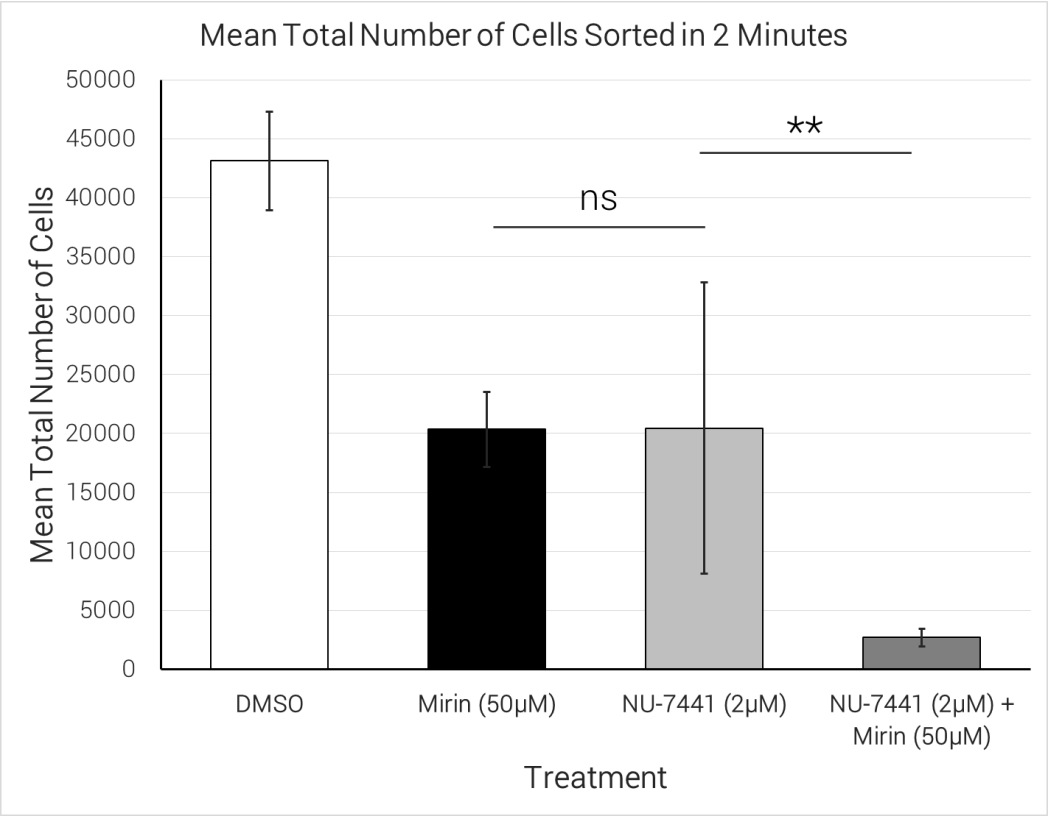


Figure 3.37: Co-treatment with the small molecules Mirin and NU-7441 is cytotoxic to MEFs at the concentrations used, 50μM and 2μM respectively. There is a synergetic effect on cytotoxicity when combining the two small molecules. Bars indicate the mean number of cells from each condition which passed through the BD LSRFortessa in 120 seconds. Error bars represent the standard deviation of three biological replicates and stars indicate the statistical significance from a Student's t-test. NB. There was a statistically significant difference between DMSO, Mirin and NU-7441 + Mirin ($p < 0.01$). The difference between NU-7441 and other conditions was not statistically significant.

3.20 Conclusions and Potential for Future Use of the *mTmG* System *in vivo*

In the course of this chapter, I have shown that the *mTmG* fluorescent reporter system provides a reliable, rapid visual read-out of CRISPR-mediated editing which can be quantified both by FC/FACS and by image analysis. It is amenable to several different transfection methodologies and can also be used to probe the effect of small molecules on editing and repair within primary mouse cells, including motile-ciliated ependymal cells which are comparable to cells affected in PCD. Furthermore, preliminary work carried out by our group has shown that this system is functional *in vivo* and could be brought forward to test gene repair in embryos and live mice. Pronuclear injections in mouse embryos, using the same reagents as these cell studies, were performed by Margaret Keighren and cultured onto blastocysts. Briefly, pX330 reagents gave low-level, mosaic editing in blastocysts, whereas RNPs provided efficient and uniform editing, including wholly HDR-derived embryos (Figure 3.38), using the minicircle repair template.

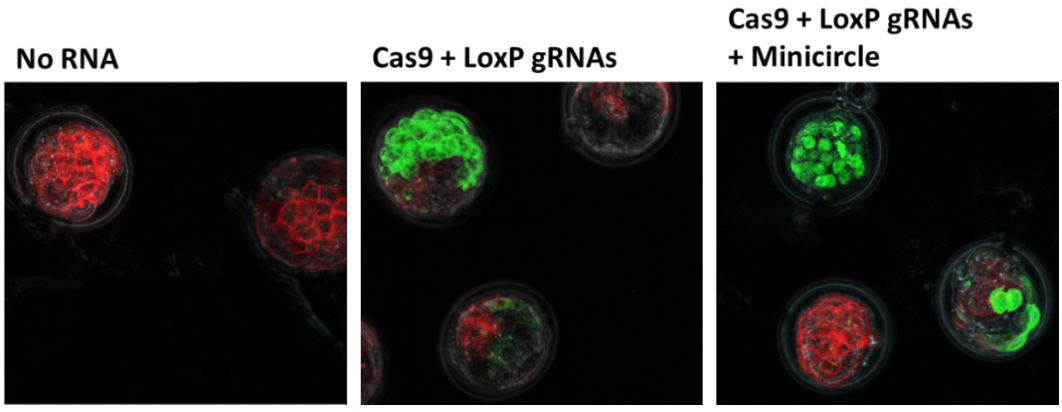


Figure 3.38: Efficient editing can be detected in the mouse blastocyst following pronuclear injection of RNPs and a minicircle repair template in fertilised eggs. Injections were performed by Margaret Keighren and images were captured by Dr. Pleasantine Mill.

In the future this system could feasibly be scaled up to perform small molecule screens for compounds which may improve repair in primary cells, both providing mechanistic insights and potentially identifying compounds which could be useful for gene therapy. While the techniques used here to measure HDR provided somewhat variable results, recent technological advancements such as the Celldiscoverer 7 and imaging-based

FC/FACS approaches should allow automated, rapid and accurate measurement of the rates of HDR across a range of treatments.

The trials which I performed with the five small molecules all proved inconclusive, despite the fact that these small molecules have all been shown to improve rates of HDR in other systems (Leahy, Golding et al. 2004, Yu, Liu et al. 2015, Song, Yang et al. 2016, Li, Zhang et al. 2017). Far from invalidating the results from our system, I believe this highlights the inherent variability in the complexes and pathways which are employed to repair genomic legions. These are likely to vary between cell types, and even within cell types depending on the cell cycle. Furthermore, there is much work to be done on understanding when and where each repair pathway is preferentially employed, depending on the size of the generated break, whether there are regions of homology beside the break, and whether a repair template is available.

3.20.1 Comparison of the *mTmG* System with other Methods Currently Used to Measure Editing and Repair

Currently, the most common methods used to measure genomic editing are based around PCR, digestion and sequencing approaches, although these methods all involve processing a mixed population of cell types and may be subject to significant technical bias. The simplest methods, based on PCR and digestion, are the Surveyor assay and restriction-fragment length polymorphism (RFLP) assays. The Surveyor assay makes use of the T7 exonuclease which cleaves mismatched bases in dsDNA; small non-homologous regions as a result of editing will therefore result in multiple bands being seen when the digestion product is run on a gel. By running an untreated PCR sample alongside the digested sample, relative intensities can be used to calculate an estimate of editing efficiencies (Figure 3.39A). However, this system cannot detect larger indels which prevent dsDNA annealing, as are created by CRISPR-mediated genome editing. Alternatively, in an RFLP assay, the presence or absence of a restriction site, as elicited by editing or repair, can be used to quantify either outcome with some accuracy (Figure 3.39B).

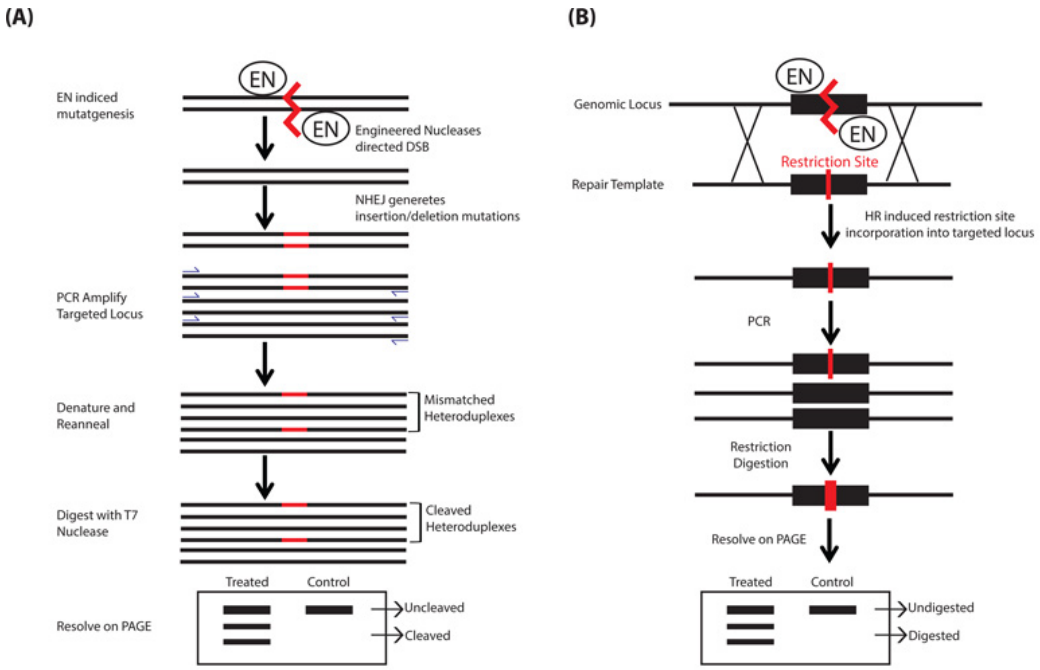


Figure 3.39: A Schematic Representation of A) a Surveyor assay and B) an RFLP assay. Taken from Hendel, Fine et al. (2015).

Recently, novel PCR-based approaches have been described which may be more accurate and flexible than Surveyor and RFLP assays. These include the use of 'droplet digital PCR', in which fluorescent probes are designed which bind specifically to the repair and wild-type sequence, and whose binding would be disrupted by editing (Figure 3.40). This assay has been shown to be able to detect HDR at a frequency of only 1 in 1000 cells (Miyaoka, Berman et al. 2016). Another alternative, termed high resolution melting analysis (HRMA), developed by Zaboikin, Zaboikina et al. (2017), is based on mismatches reducing the melting point of dsDNA, allowing sequence changes to be determined based on the melting curve. Furthermore, following PCR, there are a whole suite of potential next generation sequencing technologies which could be employed to elucidate rates of NHEJ and HDR, although these would all involve a significant outlay of time and capital, while the short reads returned may not allow confirmation of HDR if the repaired region is large. Recently described methods such as nanopore sequencing (Feng, Zhang et al. 2015) and single molecule real time sequencing (Hendel, Kildebeck et al. 2014), which provide much longer reads, may be more appropriate in these cases, although both technologies are currently in their infancy.

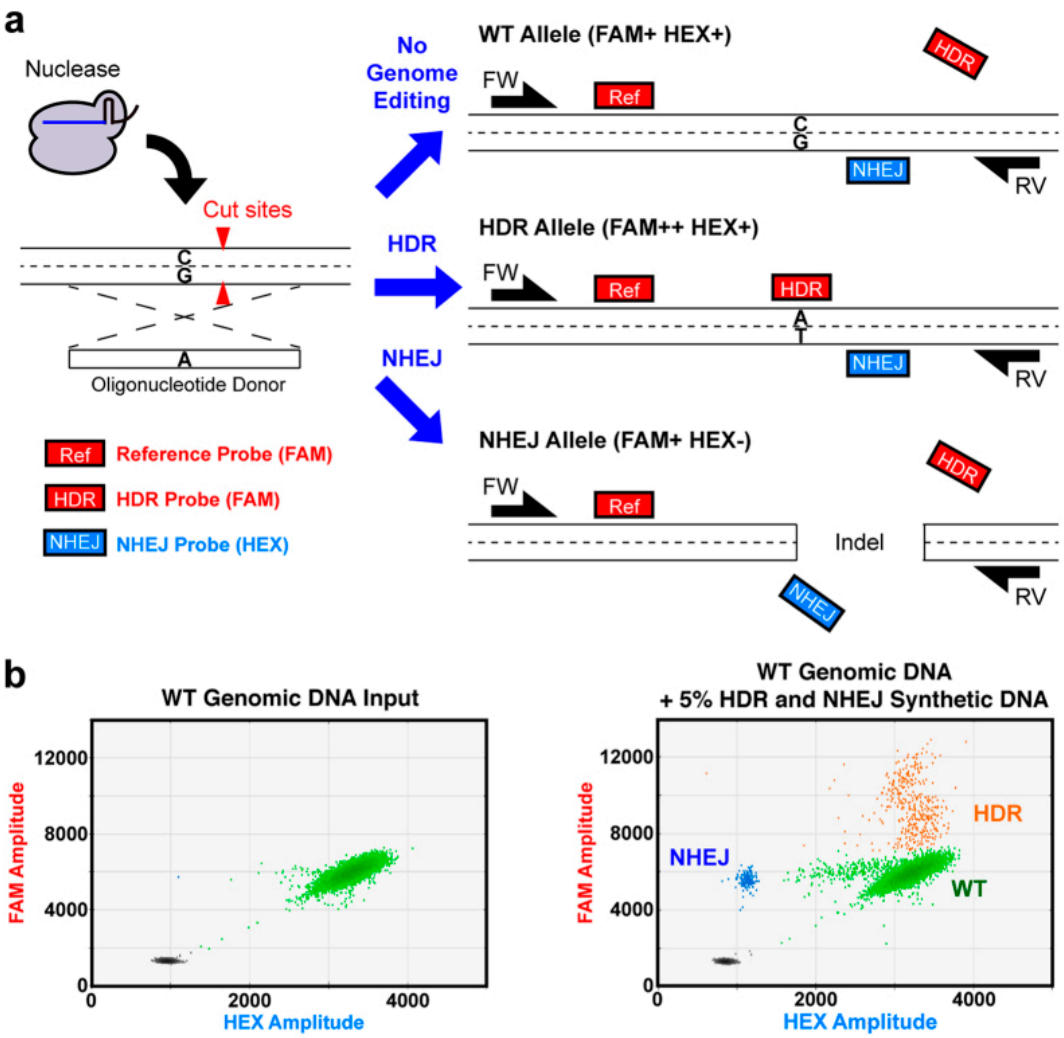


Figure 3.40: Fluorescent probes allow the identification of NHEJ and HDR in a population of edited cells upon droplet digital PCR. A) Design of the experiment and placement of probes relative to editing. B) The outcomes of editing can be visualised by fluorescence intensities. Taken from Miyaoka, Berman et al. (2016).

Alternatively, since starting this project several reporter systems have emerged which would allow a basic pre-clinical screening of compounds that may improve HDR and gene correction; mostly in cell lines such as HEK293 and U2OS. As very little is known on cell-type specific differences in DNA repair processes, this may limit the applicability of any results found, especially to therapeutically-relevant cell types.

The most established system for measuring rates of NHEJ and HDR is the Traffic Light reporter system, which was developed by Certo, Ryu et al. (2011). Similar to our system, it is based on a dual fluorescence reporter cassette (mCherry and EGFP), however they have employed a (low MOI) lentivirus to randomly integrate the cassette

into HEK293T cells. In this system, the GFP is disrupted by an I-SceI nuclease target site and the following mCherry is out-of-frame. Indels created by NHEJ will restore the mCherry reading-frame in approximately 1/3 of cases; allowing the rate of NHEJ to be quantified. However, a repair template containing an unmodified GFP sequence can also be supplied with I-SceI, allowing the rate of HDR to be quantified by the number of cells with GFP fluorescence. This system is illustrated in Figure 3.41. Using this system, the authors demonstrate that editing can be biased towards repair both by using a nickase strategy and by inhibiting DNA-PKs (as does NU-7441, Robert, Barbeau et al. (2015)). The Traffic Light reporter system has since been modified to measure the effect of both single-strand annealing (SSA) and chromatin context on repair (Kuhar, Gwiazda et al. 2014). Interestingly, they found that the rate of HDR was reduced 10-fold in a transcriptionally-inactive chromatin context, whereas the rate of NHEJ only saw a 2-fold decrease. If validated in endogenous loci, these results suggest that the status of the target locus may affect the outcome of editing and have therapeutic implications.

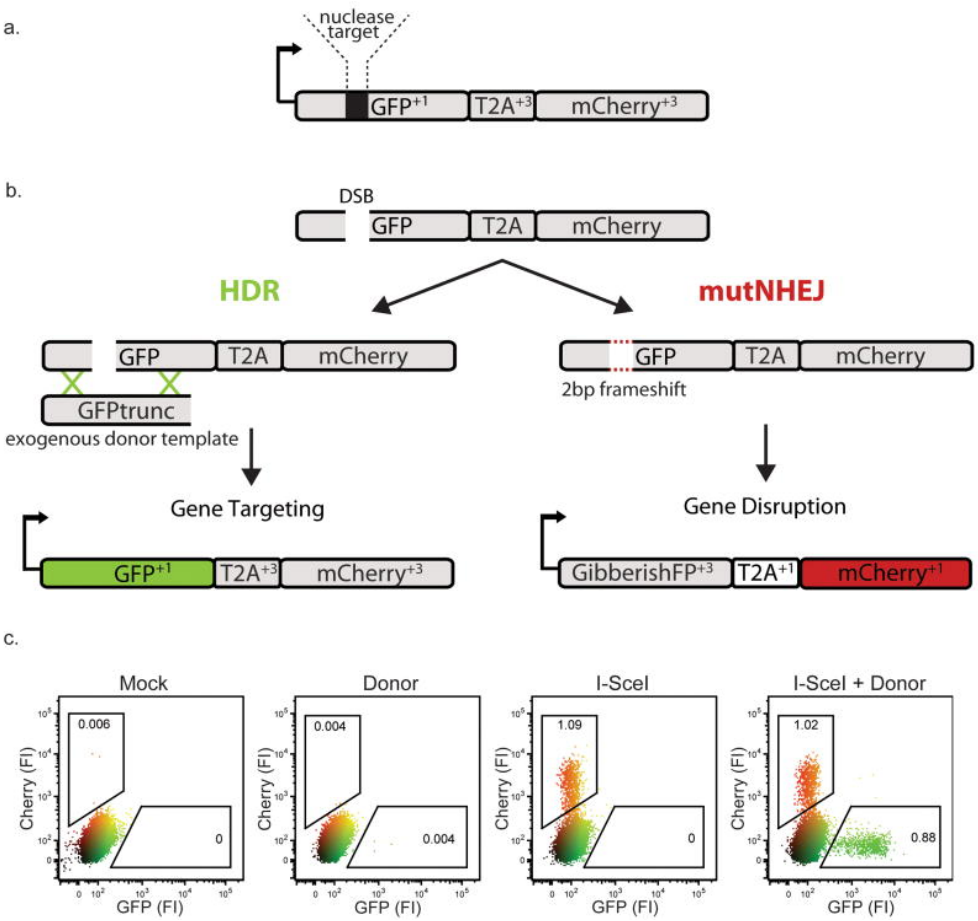


Figure 3.41: Summary of the Traffic Light reporter system. A) Schematic of the integrated fluorescence reporter cassette prior to B) editing by NHEJ or HDR. C) The resulting editing outcomes can be quantified by FC. Taken from Certo, Ryu et al. (2011).

Two other fluorescence-based reporter systems for measuring HDR have also been described. The first employs a piggyBac transposon vector containing a puromycin resistance cassette followed by a poly-A sequence and H2B-GFP reporter (Wen, Liao et al. 2017). By inserting the target sequence for any guide of interest upstream of the poly-A sequence using a BstEII site and delivering this guide together with a repair template containing Puro^{R}-T2A-H2B-GFP, the rate of HDR can be measured by the number of cells expressing H2B-GFP (Figure 3.42). It is proposed that this would be proportional to the rate of cleavage driven by this guide, although as the report by Kuhar, Gwiazda et al. (2014) indicates HDR is disproportionally inhibited by chromatin context, this may not be true at all loci. An alternative system, described by Glaser, McColl et al. (2016), simply involves the conversion of EGFP to BFP by changing three critical nucleotides in the sequence (in an 11bp window), using a

supplied repair template. In theory, this system can be employed in any cell or mouse line which expresses EGFP, and with it rates of NHEJ can be measured by loss of fluorescence whereas HDR can be detected by a change to blue fluorescence. Our *mTmG* system offers several advantages over this system, however:

a) changes in fluorescence are easier to identify than a loss of fluorescence,
b) a localisation tag allows the identification of the targeted cell types by morphology (Figure 3.14), and
c) the UV illumination used for imaging BFP results in significant phototoxicity while BFP is also relatively dim (low quantum yield and brightness).

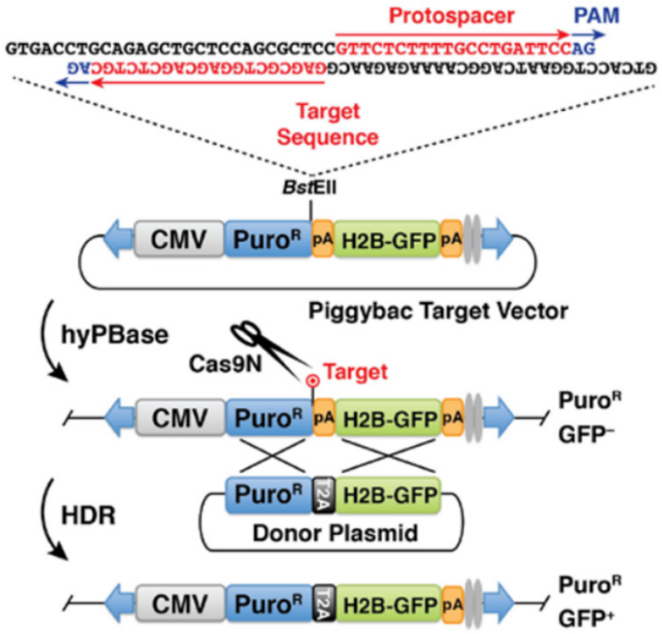


Figure 3.42: Schematic representation of the piggyBac HDR reporter system. Homology-directed repair results in the expression of H2B-GFP in the correctly targeted cells. Taken from Wen, Liao et al. (2017).

In the following chapter, I further validate our *mTmG* reporter system by performing next generation sequencing on non-edited *mTmG* cells, and all our edited sub-types. This sequencing reveals that the spectrally-distinct sub-populations we observe following transfection are indeed reflective of the outcomes we expect at a sequence-based level, providing further evidence that our system is a valid tool for measuring, and optimising, rates of CRISPR-mediated genome editing in primary cells and *in vivo*.

Chapter 4 Sequencing of the *mTmG* Locus

4.1 Introduction

As I have demonstrated that the *mTmG* System provides a reliable, visual read-out of NHEJ and HDR, I wanted to confirm that the changes visually observed are the result of the expected editing at a sequence-based level. In order to survey the full range of editing events we are seeing in each of our spectrally distinct sub-populations, as well as to hopefully detect evidence of rare HDR events, it was decided to use a next generation sequencing approach. Next generation sequencing (NGS) should provide several thousand-fold coverage in each of our populations, allowing us to detect and quantitate even very rare events, while being much less labour-intensive than other manual sequencing approaches, such as TOPO-cloning.

For this work, I took advantage of the services offered on-site by the Edinburgh Wellcome Trust Clinical Research Facility (WTCRF) in designing and carrying out NGS sequencing experiments on two complementary platforms. The first, the Ion Torrent Personal Genome Machine (PGM) system (Thermo Fisher), offers long reads, up to 400bp, as well as high coverage and a low error rate (1.71%, (Quail, Smith et al. 2012)). This makes it highly suited for an experiment such as ours in which we want to detect editing in a repetitive locus. For longer reads across our 3kb locus, I also decided to test a novel sequencing technology, the Oxford Nanopore MinION. This platform can theoretically provide reads several thousand basepairs long, and so is highly suited to detect large deletions and insertions as I would expect in this experiment. The limitation of this system is its high error rate, around 10% in current iterations, and thus it is not the ideal system for detecting small indels and basepair changes. Therefore, I viewed these platforms as complementary; the Ion Torrent is able to provide us with accurate details of small edits at either cut site, while the MinION should provide evidence of larger deletions and insertions, such as of the H2B sequence.

Following sequencing, I used two approaches to analyse changes at our locus. The first was variant calling, in which the resulting sequencing reads are mapped back onto the reference sequence and then any variants are detected, quantified and potentially

assigned a probability score. These variants can be viewed visually on a genome browser or called by a variant calling programme. This approach is generally more suitable when the variable regions in the supplied sequences are relatively small and few in number, as the presence of a very large indel or several smaller changes may result in the failure of the read to align back to the reference. Furthermore, one must consider that variant calling programmes are currently designed to detect discrete mutations, for example a +AG insertion at a single position, such as observed in disease mutation genetics, as opposed to the large genomic variation resulting from a CRISPR/Cas9 cut site.

An alternative approach to variant calling, which may be particularly useful in our case, is *de novo* genome assembly. During *de novo* genome assembly, the reads from the sequencing experiment are aligned using their overhangs to attempt to form a new 'reference' sequence or sequences. The number of reads which go to form each new reference sequence can then be quantified. This approach is useful for detecting large deletions, insertions and rearrangements, although it will ignore small, infrequent variants when forming a consensus sequence. However, if required, one could use the output reference sequences from a *de novo* genome assembly for further variant calling on the initial data, and potentially recover these smaller, less common variations. There are also a number of stand-alone tools designed to analyse NGS results from CRISPR experiments, such as CRISPResso (Pinello, Canver et al. 2016), however due to the unique nature of the editing expected in the *mTmG* reporter system, with two cut sites, a large deletion and potentially also a large insertion present, these 'off-the-shelf' analysis suites cannot currently be applied to our data.

During this chapter I will detail the approaches I used to sequence the *mTmG* locus, including the initial design, and then refinement of the experiment, the laboratory work carried out prior to sequencing, and the various methods used to analyse the sequencing data. I show that while there are several challenges in sequencing this locus, and analysing the resulting data, I can demonstrate that the four populations we observe following transfection, that being Tomato-positive, GFP-positive, double-positive and double-negative, are reflective of the edited sequences in these cells. Further work is required, however, to accurately detect and quantify cells containing the H2B repair sequence, for reasons I will discuss.

4.1.1 Experimental Design

In order to sequence across the *mTmG* locus, I designed four sets of primers to attempt to capture all of the possible changes we may see following editing and repair in our system. These four sets of primers are listed in Section 2.5.1, and have been plotted visually onto the *mTmG* locus in Figure 4.1:

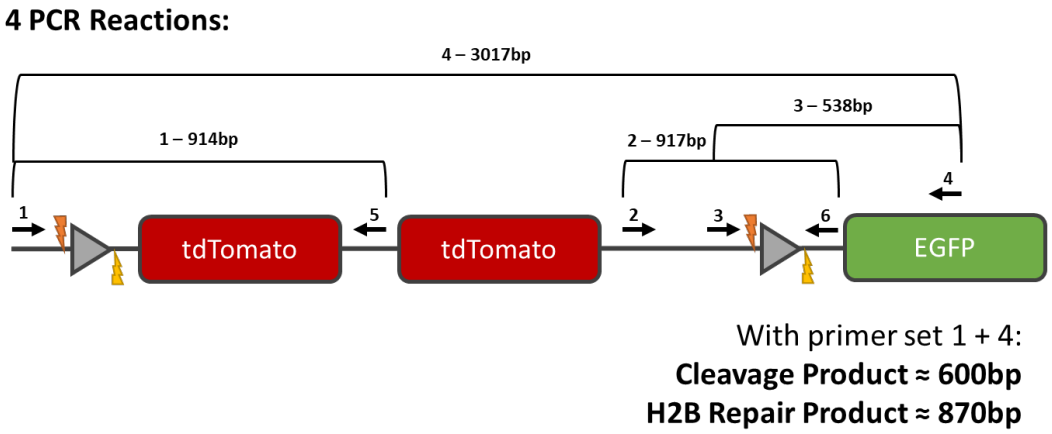


Figure 4.1: Location of the sequencing primers on the *mTmG* locus. Primers are indicated by arrows with their respective numbers above. PCR products and their expected sizes are noted above primer pairs with brackets. The loxP sites are indicated with grey triangles, and the guide target sites either side are indicated with lightning bolt symbols. The sizes of the cleavage and repair products are approximate due to the variable sizes of indels generated by Cas9. NB. Product sizes are based on the sequencing of the locus in Section 4.3. Diagram not to scale.

Primer choice was limited due to the repetitive nature of the *mTmG* locus, resulting in the use of primers with relatively large products which will need to be sheared prior to Ion Torrent sequencing. In design, PCR product 1 was chosen to allow me to quantify NHEJ at the first loxP site which hasn't resulted in the excision of Tomato, while PCR products 2 and 3 should allow me to do the same at the second loxP site. PCR product 4 can then be used to detect larger changes, such as the deletion of the region between the loxP sites to produce the 'Cleavage Product' and the incorporation of the H2B sequence to form the 'H2B Repair Product'. Neither should be detected in the other PCR products due to the deletion of the binding sites for primers 2, 3 and 5.

Using these primers, I performed PCR amplifications with the proof-reading polymerase, Phusion Hot Start II (Section 2.6). PCR amplifications were carried out on DNA extracts from the edited sub-populations (Tomato-positive, GFP-positive,

double-positive and double-negative), which were all isolated by FACS. Following PCR amplification, PCR products were purified (Section 2.4.2) and submitted for pre-processing and sequencing by the WTCRF (Section 2.18 for details).

4.2 Initial Ion Torrent Run

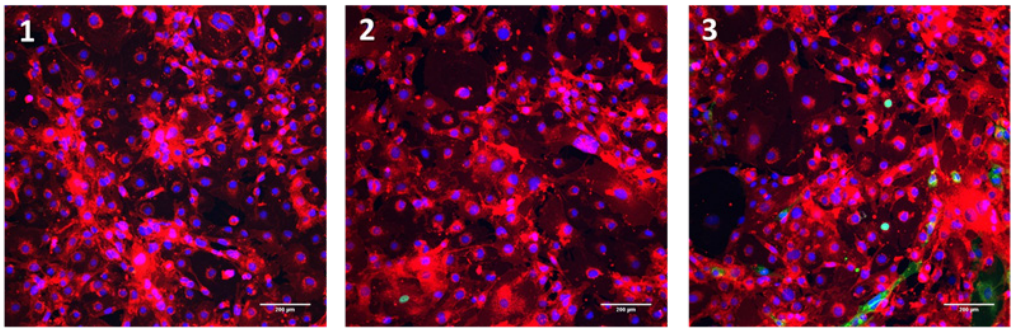
4.2.1 CRISPR Genome Editing at the *mTmG* Locus Results in Four Spectrally Distinct Sub-Populations

For the initial sequencing run on the Ion Torrent, I transfected immortalised heterozygous *mTmG* MEFs with the LoxP pX330 plasmids and the H2Bv2 repair plasmid, which had been shown to drive repair (Section 3.12). Three transfections were carried out on the Neon system (Section 2.12.1.1) using the 10µl electroporation tips and $1.5x10^5$ cells per well (3 reactions):

Well	Cell Number	Cas9 Plasmid(s)	Cas9 Plasmid DNA (ng)	Repair Plasmid	Repair DNA (ng)
1	$1.5x10^5$	N/A	0	N/A	0
2	$1.5x10^5$	N/A	0	H2Bv2	1500
3	$1.5x10^5$	LoxP#1 pX330, LoxP#2 pX330	1500 (725/plasmid)	H2Bv2	1500

Ninety-six hours post-transfection, each well was split 1 in 2. Twenty-four hours later, each well was imaged seven times in random positions at 15x on a Nikon A1R confocal. These images were then quantified using the Cell Counter and Nucleus Counter plugins in Fiji (Section 2.16.1), which revealed that the transfections had been successful and that the transient expression from the H2Bv2 plasmid had mostly dissipated (Figure 4.2A, B). Based on this, each well was then dissociated and sorted on a BD FACSAria twenty-four hours later. As expected, the vast majority of cells in the control wells were Tomato-positive, whereas the cells in well 3 split into the four populations seen previously; Tomato-positive, GFP-positive, double-positive and double-negative (Figure 4.22C).

A



B

Well	GFP-positive Cells (%)	GFP-positive Nuclei (%)	Total Cells	Treatment
1	0.00	0.00	2461	No DNA
2	0.00	0.05	1894	H2Bv2 Only
3	5.91	0.63	1894	LoxP#1 + LoxP#2 + H2Bv2

C

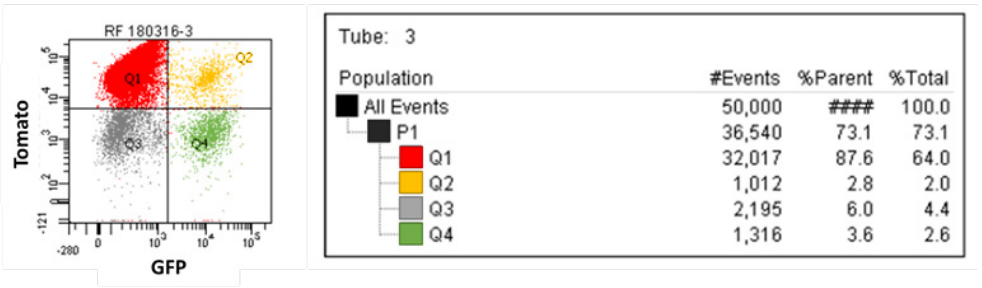


Figure 4.2: Transfection and isolation of genome edited spectrally distinct sub-populations of *mTmG* MEFs for NGS. A) Representative confocal images of each well taken at 15x, 96 hours post-transfection. Scale bar 200µm. B) The images from each well were quantified using Cell Counter and Nucleus Counter in Fiji in order to estimate rates of NHEJ and HDR. C) An example FACS plot from well 3 in which the cells can be seen to form four populations; Q1 – Tomato-positive, Q2 – double-positive, Q3 – double-negative and Q4 – GFP-positive. NB. Results are only shown for the final 50,000 cells sorted. Other FACS plots included on attached CD in file 'Figure 4_2'.

4.2.2 FACS Is Able to Sort Edited Cells into Their Sub-Populations with Reasonable Accuracy

Following FACS, the two control populations and the four edited populations were returned to a 6-well plate and allowed to reach confluency. At this point, I imaged each well five times at 20x on a Nikon A1R confocal. Imaging revealed that there were no GFP-positive nuclei remaining in the H2Bv2 only control, as expected. As previously observed, there was significant contamination of the Tomato-positive population into the other wells (Figure 4.3). These images were quantified using the Cell Counter and

Nucleus Counter plugins, which indicated that ~10% of the GFP-positive sort comprised Tomato-positive cells along with ~3% of the double-negative sort. Later experiments have indicated that this appears to be unavoidable, see for example Section 3.14. This cross-over appears to be due to the abundance of the Tomato-positive population, potentially combined with the adherent nature of MEFs.

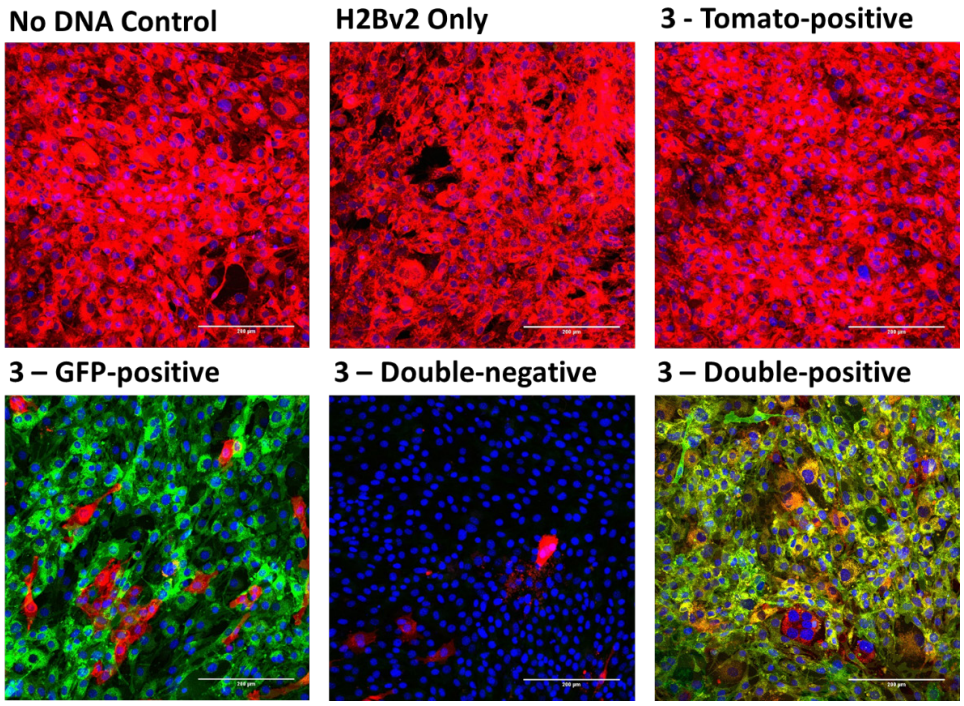


Figure 4.3: *mTmG* MEFs maintain their edited states over time, although there is an error rate of up to 10% in FACS. Representative confocal images captured at 20x, 16 days post-FACS. Scale bar 200µM.

4.2.3 PCR Optimisation for Sequencing

Following imaging, I extracted DNA from each well using Quick Lysis Buffer (Section 2.4.1). DNA extractions were then quantified using a Nanodrop spectrophotometer. Dilutions of each DNA extraction were therefore made, at a concentration of 100ng/µl in NF-H_2O. After several rounds of PCR optimisation, each primer set produced a single band of the predicted size (in samples where only one band was expected). The inclusion of betaine in the reaction mixes was found to be necessary to achieve amplification, indicating there may be strong secondary structures present. Furthermore, annealing temperatures had to be raised several degrees above their expected values to prevent non-specific binding. The final reaction compositions used for each primer set are detailed in Section 2.6.1, while the general

cycling parameters are detailed in Section 2.6.2. Following PCR, each amplification was purified using a QIAquick PCR Purification Kit (Section 2.4.2) and 5µl of each purified product was electrophoresed on a 2% agarose gel. This indicated that bands of the expected sizes had been produced for each PCR product (Figure 4.4). Each product was named with a letter representing the DNA template followed by the number of the PCR reaction:

A - NF-H_2O
B - Cas9 Only Control
C - H2Bv2 Only Control
D - Tomato-positive
E - GFP-positive
F - Double-positive
G - Double-negative

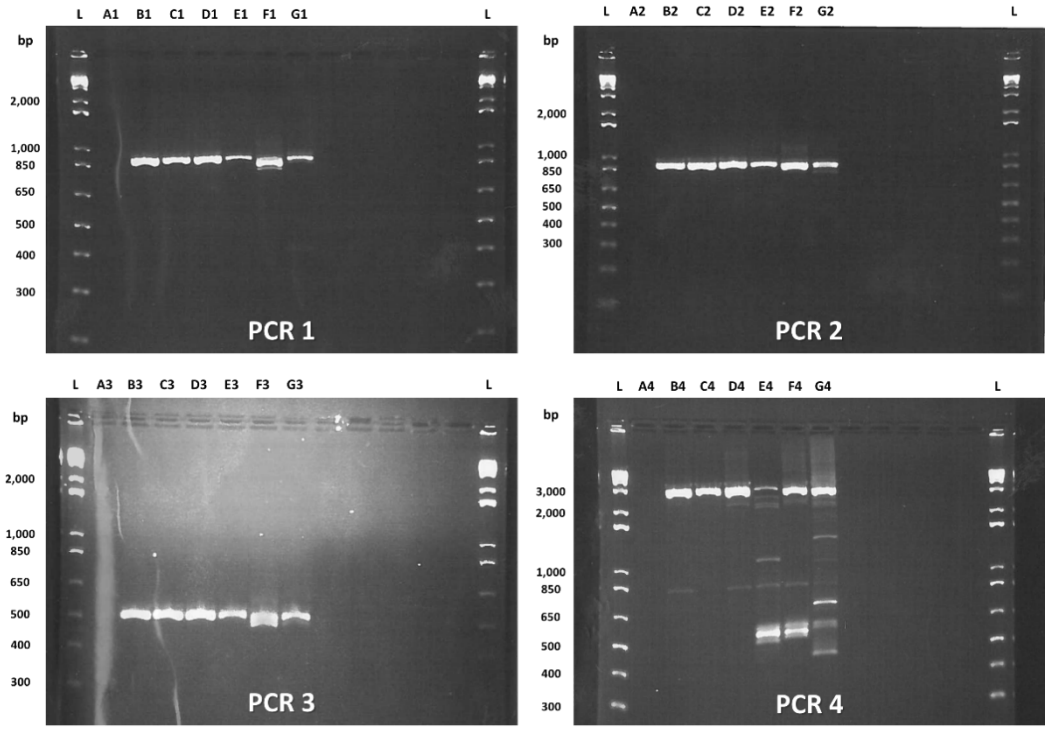
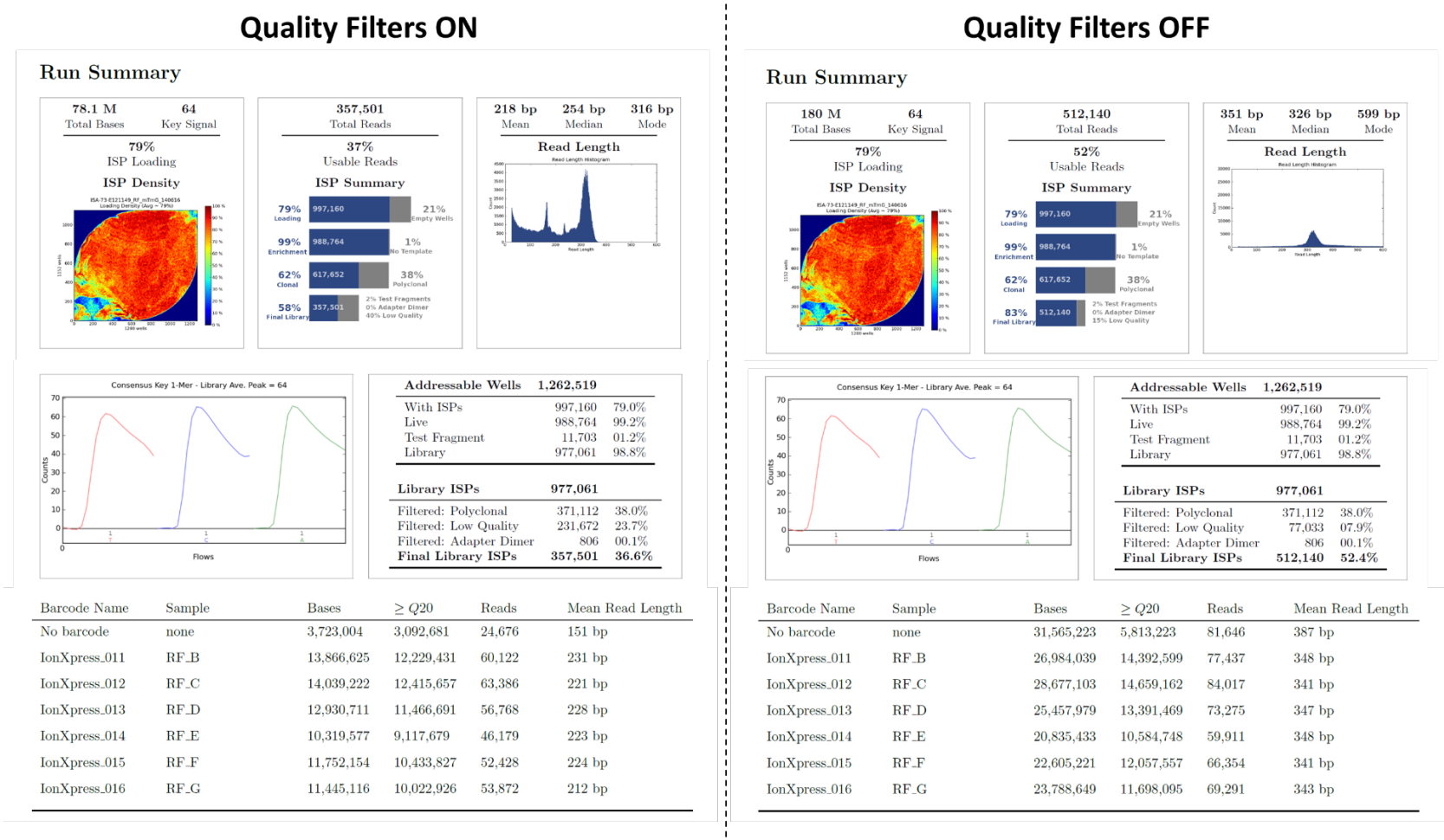


Figure 4.4: Purified PCR products for Ion Torrent sequencing. Products were electrophoresed on a 2% agarose gel. L = 1kb Plus DNA Ladder (Thermo Fisher).

4.2.4 Sequencing Design and Ion Torrent Run Reports

Following PCR purification, each sample was transferred to the WTCRF for preparation and sequencing. Samples were prepared as detailed in Section 2.18. All PCR products were sheared to a mean length of 350bp and then six different barcodes were added to allow us to separate reads from each DNA template *in silico*. Run reports were automatically generated by the Ion Torrent software upon completion of sequencing. The initial run report indicated that the mean read length was significantly shorter than we were expecting, at 218bp (Figure 4.5, quality filters on, top right box). Furthermore, ~24% of the reads were being discarded or trimmed due to their quality scores (Figure 4.5, quality filters on, middle right box). It was decided to reanalyse the data with the quality filter relaxed, which increased the mean read length back up to its expected value at ~350bp (Figure 4.5, quality filters off, top right box). Therefore, this indicates many of the reads substantially declined in quality towards the 3' end and as such have been trimmed by the software. As low-quality bases may complicate any analysis, it was decided to initially continue with the reads produced with the more stringent quality filter. Based on the number of reads per barcode, this should still leave several thousand reads per PCR product. The resulting reads, in fastq format, were aligned and analysed using the pipeline detailed in Section 2.19.1 to the expected (reference) sequences, along with the predicted cleavage product and H2B repair product.



Barcode Name	Sample	Bases	≥ Q20	Reads	Mean Read Length
No barcode	none	3,723,004	3,092,681	24,676	151 bp
IonXpress_011	RF_B	13,866,625	12,229,431	60,122	231 bp
IonXpress_012	RF_C	14,039,222	12,415,657	63,386	221 bp
IonXpress_013	RF_D	12,930,711	11,466,691	56,768	228 bp
IonXpress_014	RF_E	10,319,577	9,117,679	46,179	223 bp
IonXpress_015	RF_F	11,752,154	10,433,827	52,428	224 bp
IonXpress_016	RF_G	11,445,116	10,022,926	53,872	212 bp

Barcode Name	Sample	Bases	≥ Q20	Reads	Mean Read Length
No barcode	none	31,565,223	5,813,223	81,646	387 bp
IonXpress_011	RF_B	26,984,039	14,392,599	77,437	348 bp
IonXpress_012	RF_C	28,677,103	14,659,162	84,017	341 bp
IonXpress_013	RF_D	25,457,979	13,391,469	73,275	347 bp
IonXpress_014	RF_E	20,835,433	10,584,748	59,911	348 bp
IonXpress_015	RF_F	22,605,221	12,057,557	66,354	341 bp
IonXpress_016	RF_G	23,788,649	11,698,095	69,291	343 bp

Figure 4.5: Ion Torrent run reports indicate read quality is low after ~220bp (c.f. mean read length, top right, with quality filters on and off). However, after quality filtering, there are still a high number of reads returned per barcode (bottom table) for future analysis.

4.2.5 Sequencing Results

4.2.5.1 Alignments and Variant Calling Prove Inconclusive Due to the Highly Repetitive Nature of the *mTmG* Locus

Following alignment of the reads to the reference sequences, the mapping statistics were consulted. These revealed that the vast majority of the reads were mapped successfully back onto the references (Figure 4.6A), however there were a significant number of reads which could not be mapped uniquely to a single position (Figure 4.6B). Taking PCR product 4 to demonstrate, the aligned reads were viewed using the Integrative Genomics Viewer (IGV, Figure 4.7). On this platform, grey bars indicate uniquely aligned reads, while white bars indicate reads which aligned to multiple positions and black bars indicate deletions. Furthermore, vertical purple bars indicate insertions while coloured bars indicate single nucleotide variants (SNVs). Viewing these alignments, it becomes clear that the software, with ~220bp reads, is unable to definitively separate reads which align around either of the two loxP sites, most likely due to the high sequence similarity and the presence of indels. Therefore, it is difficult to say whether the variants seen at these sites are real or simply errors of alignment.

Aligning the samples against the expected cleavage product also produces somewhat ambiguous results. Many reads from the controls also align to this sequence, including over the loxP site, although there is a large increase (~20%) in the number of reads which align in the edited sub-populations (Figure 4.8A and 4.9). Furthermore, unexpectedly, there is no discernible enrichment in the number of reads which align to the H2B sequence between the edited sub-populations (Figure 4.8B), although very few reads align to this sequence from any of the samples (see coverage plots in Figure 4.10).

Considering these results, it was clear that we had underestimated the impact the repetitive nature of the *mTmG* locus would have on read alignment, although the problem may have been compounded by the unexpectedly short reads we received from sequencing. Based on this, it was decided to redesign and repeat the sequencing experiment.

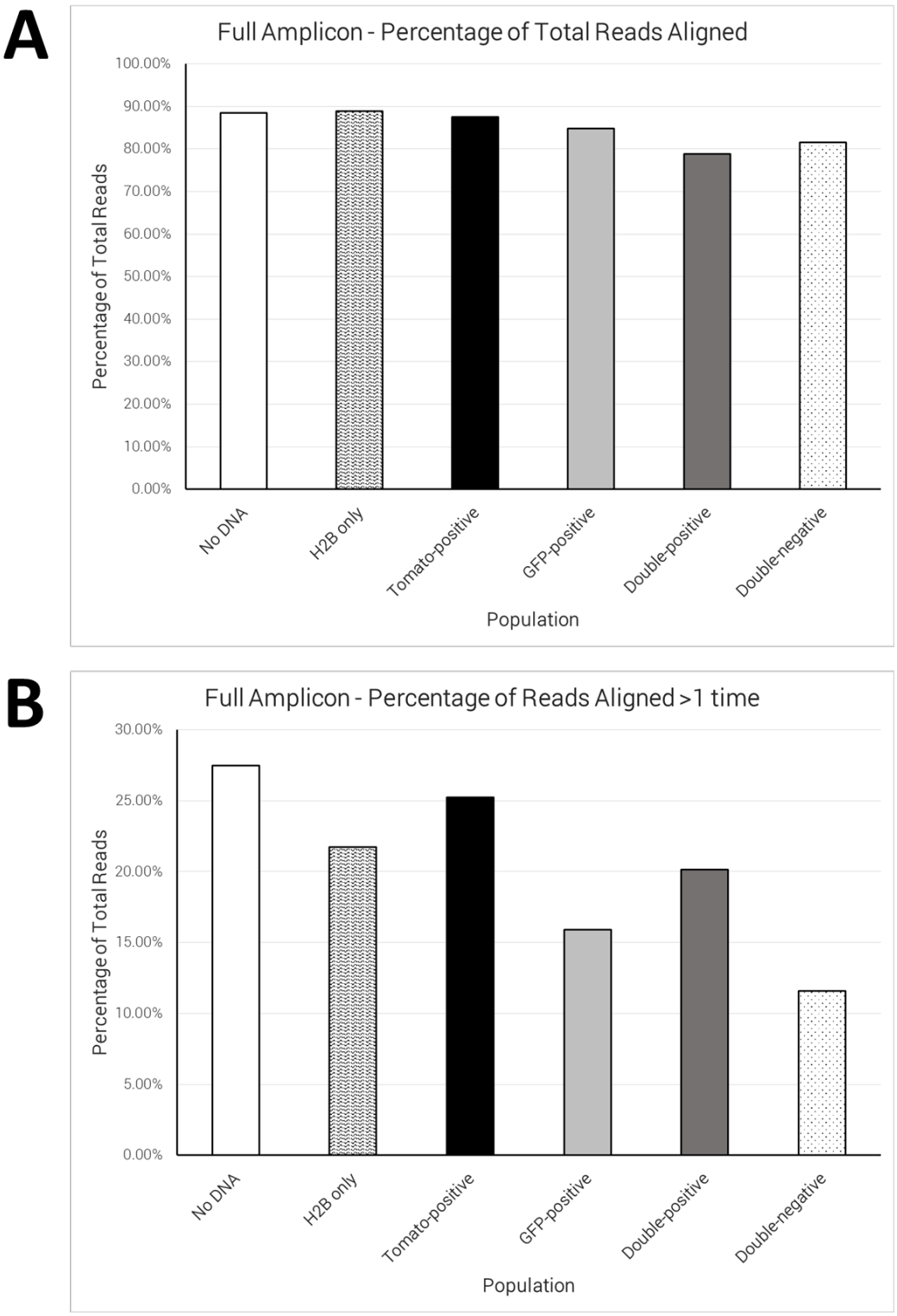


Figure 4.6: Mapping statistics following the alignment of Ion Torrent reads from each sample against the expected sequence for PCR product 4. A) The total percentage of reads which align to the reference sequence. B) The percentage of reads which can be aligned equally well to multiple positions in the reference sequence.

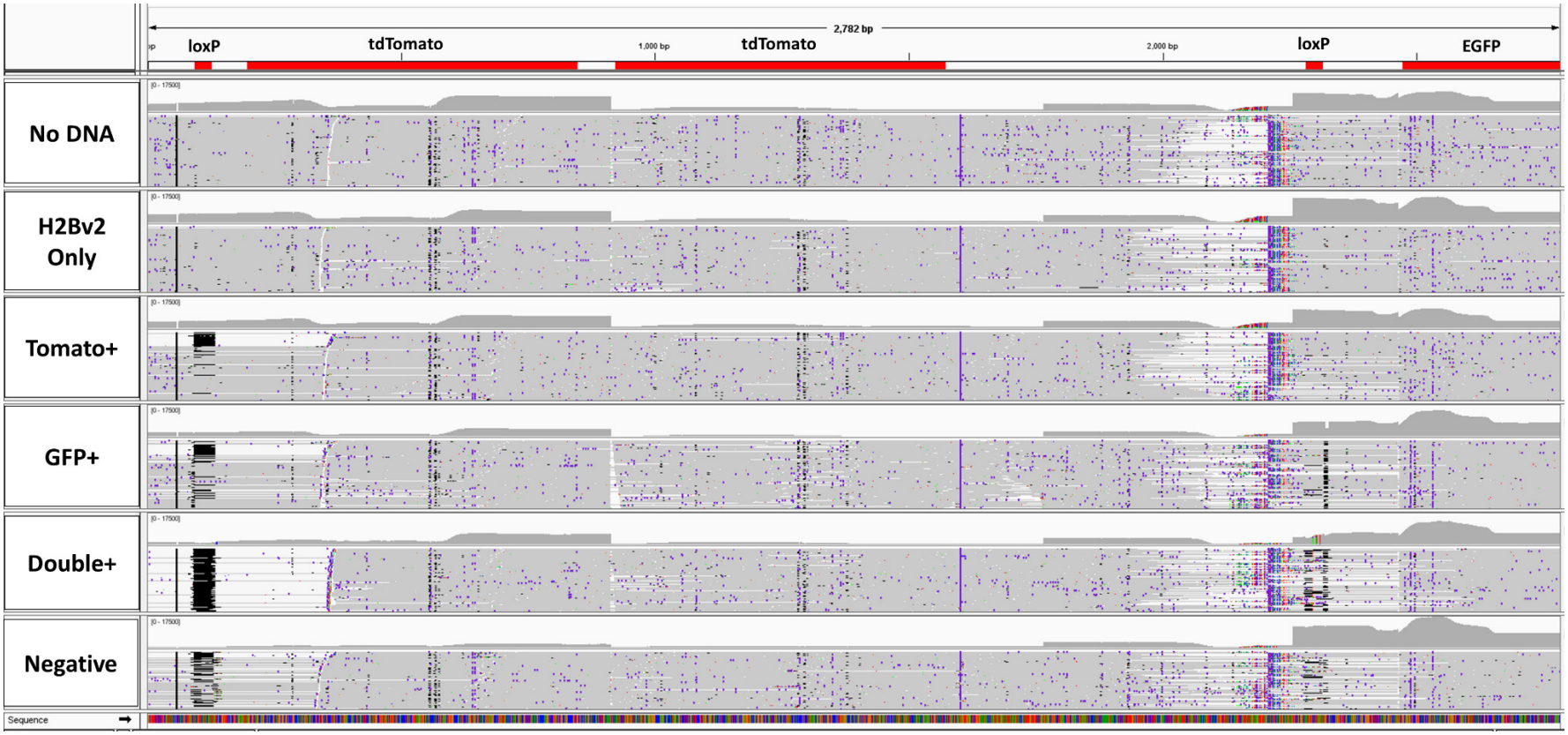


Figure 4.7: With the current data it is not possible to unambiguously align reads to either loxP site. A screen capture from the IGV browser is shown of the alignments produced for each sample against the expected sequence for PCR product 4. Here, grey bars indicate uniquely aligned reads, white bars reads which aligned to multiple positions, black bars deletions and purple bars insertions. Furthermore, SNVs are indicated with coloured bars. A coverage plot, representing the number of reads which have aligned at each position, is displayed above the alignments. Regions of interest have then been highlighted with red bars.

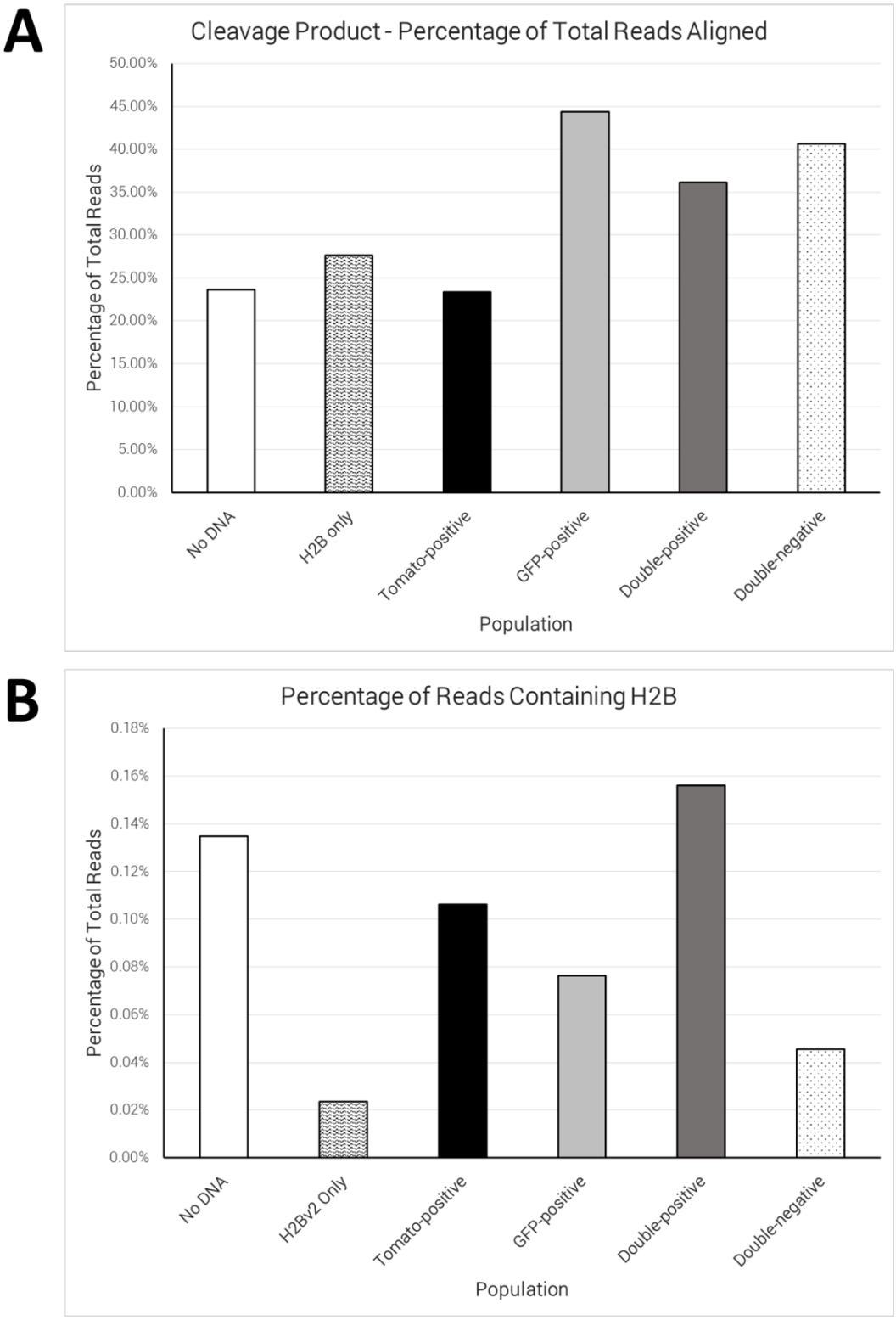


Figure 4.8: Mapping statistics showing the percentage of reads from each sample which aligned to A) the cleavage product and B) the H2B sequence. H2B is highlighted with a red bar in Figure 4.10.

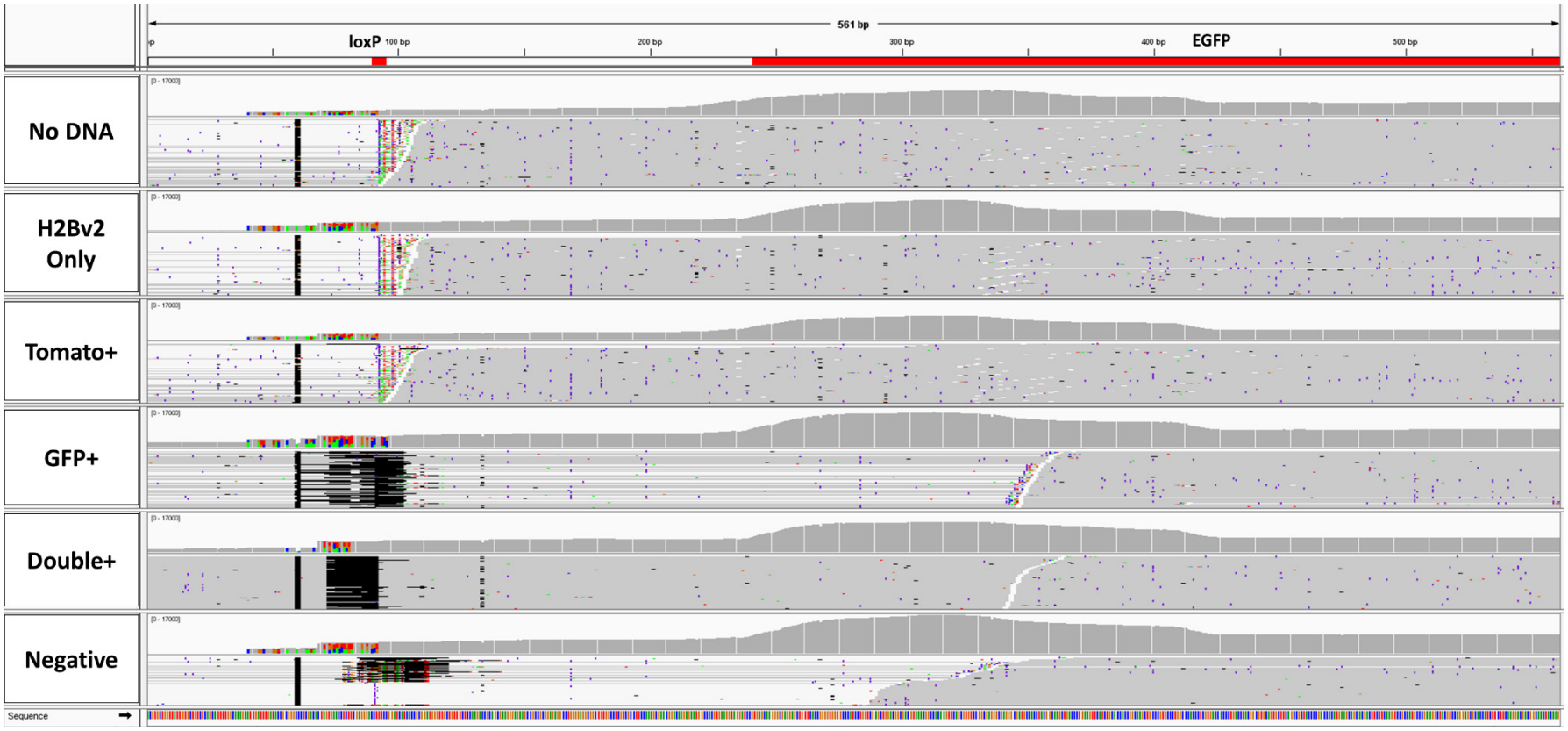


Figure 4.9: Screen capture from the IGV browser showing the alignments of Ion Torrent reads from each sample against the predicted sequence for the cleavage product, with the associated coverage plots above. The loxP site and EGFP are highlighted with red bars.

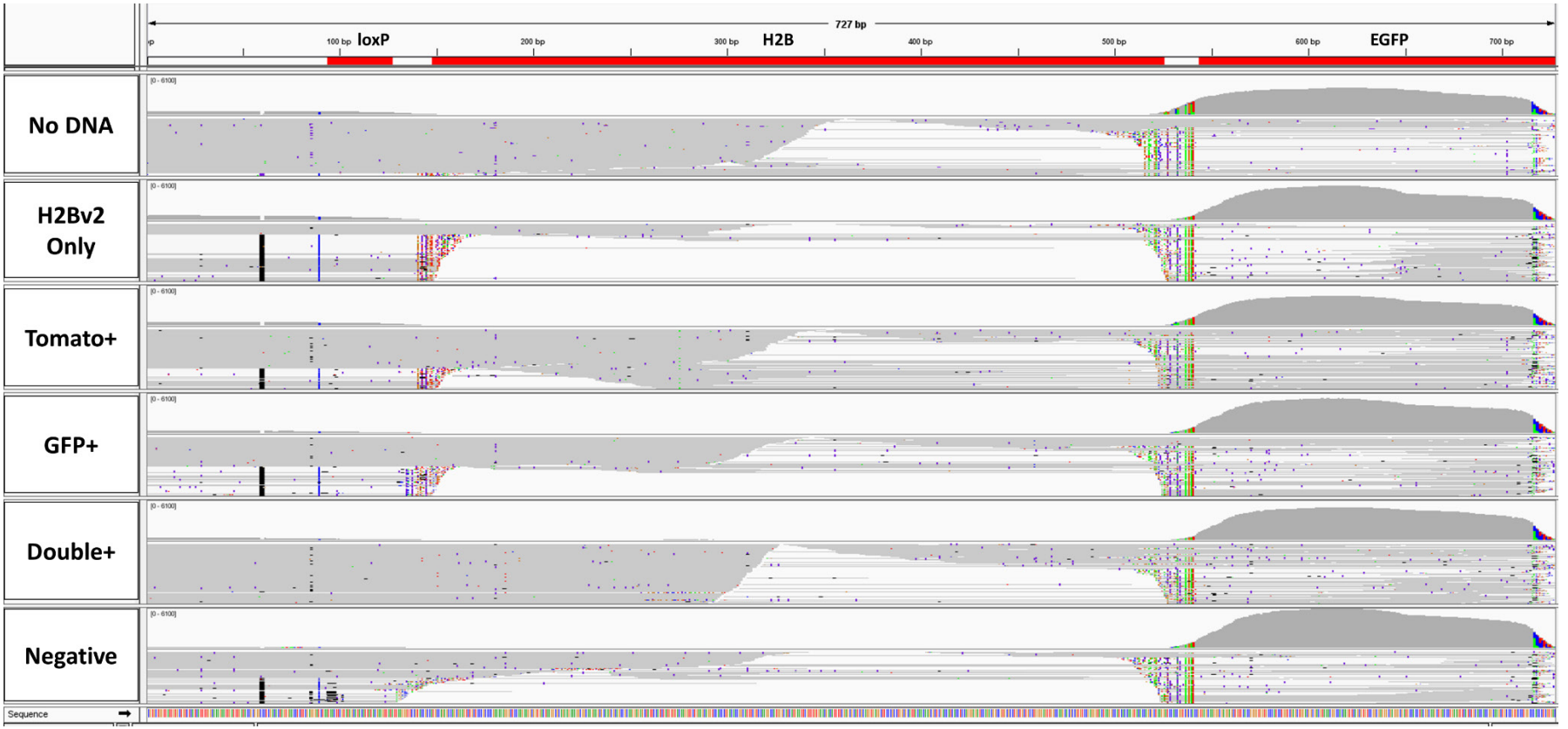


Figure 4.10: Screen capture from the IGV browser showing the alignments of Ion Torrent reads from each sample against the predicted sequence for the H2B repair product, with the associated coverage plots above. The loxP site, H2B sequence and EGFP are highlighted with red bars.

4.3 Sequencing Across the Locus Reveals Changes Relative to the Reference

Prior to repeating the sequencing experiment, I decided to Sanger sequence across the locus to determine whether there are any differences between the reference sequence and the sequence present in our MEFs. Several variants, which appear to be present in all samples, can clearly be seen in Figure 4.7; and include a deletion before the first loxP site, an insertion following the second Tomato dimer and a larger variable region before the second loxP site.

In order to sequence across the locus, I prepared a new PCR amplification with all four primer sets, using the Cas9 only control DNA and the same PCR conditions as previously. These PCR amplifications were visualised on a 2% agarose gel (Figure 4.11A), before being submitted for sequencing by technical services. The sequences were aligned against the current reference using SnapGene. This revealed, as expected, that there were two differences between the reference sequence and the sequence in our MEFs (Figure 4.11B), as well as one 218bp fragment which was missing from the sequence we were using for alignment. This 218bp fragment may have reduced the number of reads from PCR product 2 which were able to align in Figure 4.7, although due to the trimming of reads into ~220bp fragments and the coverage from PCR product 3 over the second loxP site, it shouldn't have significantly affected the overall results. Furthermore, none of the detected changes would significantly affect the functionality of our guides or repair template. These changes were added to our reference sequences for future work.

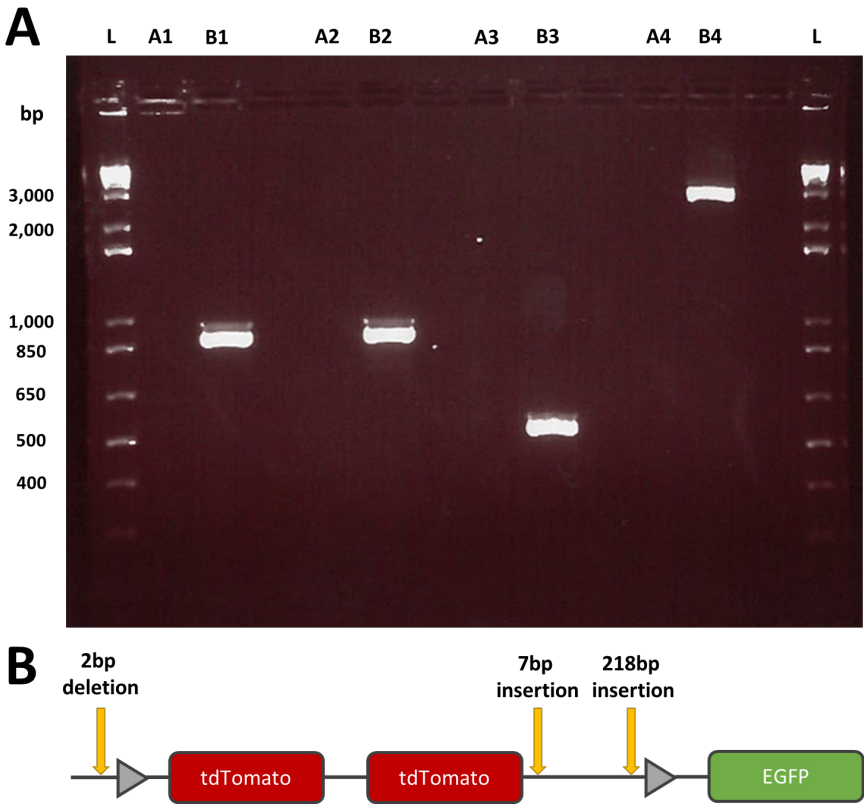


Figure 4.11: Sanger sequencing reveals two differences between the reference sequence and the sequence in the *mTmG* MEFs. A) PCR products generated using the Cas9 Only control template DNA were electrophoresed on a 2% agarose gel. L= 1kb Plus DNA Ladder (Thermo Fisher). For expected product sizes see Figure 4.1. B) Cartoon showing the size and location of the three detected changes on the *mTmG* locus.

4.4 Experimental Re-design – Barcoded Primers and A Potential Application for Nanopore Sequencing

In order to circumvent the difficulties in aligning reads back to the repetitive reference, I decided to use a more sophisticated design for sequencing, in which all of the PCR products are individually barcoded, allowing separation *in silico*. This should significantly reduce the noise seen due to incorrect alignments. Furthermore, it was decided to sequence the products on both the original Ion Torrent PGM system, and the novel Oxford Nanopore MinION system. As explained in the introduction, the longer reads from the MinION should cover the entire length of the PCR products, avoiding the issues with aligning shorter reads. However, the high error rate of the MinION makes it unattractive as a sequencing approach to quantify editing on its own, hence why it is necessary to have the Ion Torrent for comparison. For this sequencing

experiment we also decided to take advantage of several newly developed reagents which had demonstrated high efficiencies in recent experiments; that being the use of RNP complexes to direct cleavage, and a minicircle repair template to drive higher rates of HDR. These are further investigated in Section 3.16.2. Furthermore, as we tend to see a large number of multi-nucleated cells following immortalisation (data not shown), it was decided to return to using non-immortalised MEFs since duplication of the locus may complicate the sequencing results.

4.4.1 High Rates of Editing Are Seen Following RNP Transfection

Early passage non-immortalised heterozygous *mTmG* MEFs (P3) were transfected using the Neon system (Section 2.12.1.3), however, for greater numbers I used 100µl tips with $3x10^5$ cells per reaction/well, scaling up all reagents accordingly. Two wells of a 6-well plate were prepared:

Well	Cell Number	LoxP#1 RNP (ng)	LoxP#2 RNP (ng)	Minicircle (ng)	Cas9 (µg)
1	$3x10^5$	0	0	0	6
2	$3x10^5$	720	720	3000	6

Ninety-six hours later, I imaged each well at 20x on a Dragonfly spinning disk confocal (Andor). Images revealed a high transfection efficiency, and several cells with GFP-positive nuclei could be seen (not quantified, Figure 4.12A), therefore I decided to continue with FACS. Twenty-four hours later, each well was dissociated and sorted on a BD FACSAria, separating each well into the four populations as before but with more constricted gates. This time, a small negative population was also captured from the control for inspection (Figure 4.12B). The rate of editing was found to be high in well 2 (~26%) and, as I had seen before, there was a low proportion of double-positive cells (1%) in the non-immortalised MEFs. FACS plots in Figure 4.13 are shown for a proportion of the cells, the final numbers sorted from well 2 are approximately:

Tomato-positive	-	400,000 cells
Negative	-	20,000 cells
GFP-positive	-	20,000 cells
Double-positive	-	3,000 cells

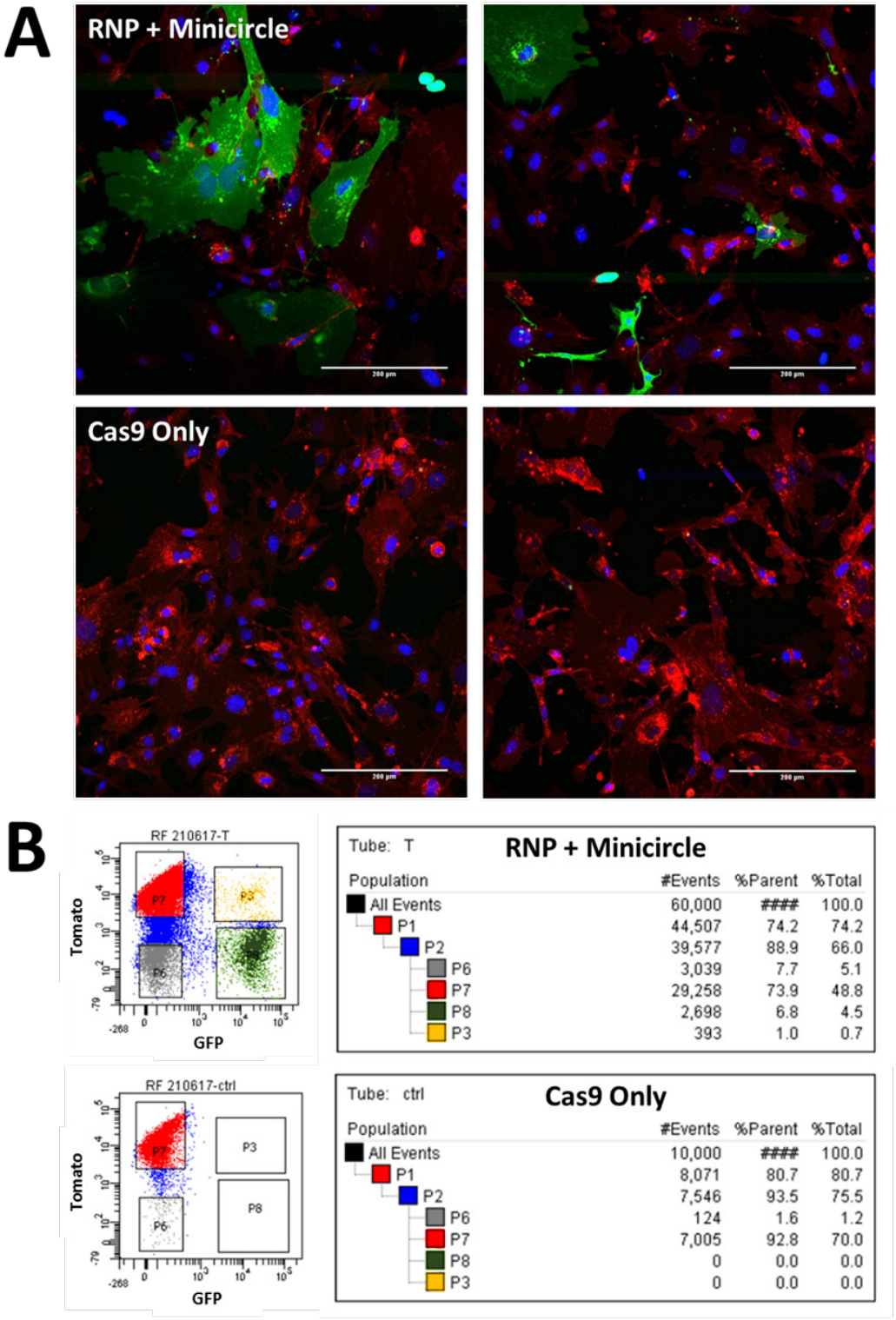


Figure 4.12: Joint transfection with RNP complexes and a minicircle repair template leads to a high rate of NHEJ and HDR in *mTmG* MEFs ahead of deep sequencing. A) 20x confocal images taken 96 hours post-transfection. Scale Bar 200µm. B) Each well was sorted by FACS, 120 hours post-transfection. FACS plots for each well are shown, along with the gates used for sorting.

Each of the sorted populations was returned to a 24-well glass-bottomed plate for outgrowth. Forty-eight hours later, the majority of the wells were reasonably confluent, and each was imaged five times on a Dragonfly confocal at 20x. Images revealed that the sort had been largely successful, although again there was cross-over of cells into other populations. Interestingly, the 'negative' population from the control well appeared to be entirely Tomato-positive, if somewhat dimmer (Figure 4.13). Furthermore, several cells with nuclear-GFP could be seen in the GFP-positive and double-positive wells (not quantified). Based on these positive results, I extracted DNA from each well, at seven days post-transfection, following the protocol in Section 2.4.1.1. DNA concentrations were reasonably low, between 10-20ng/µl, for several of the populations, due to the paucity of input cells.

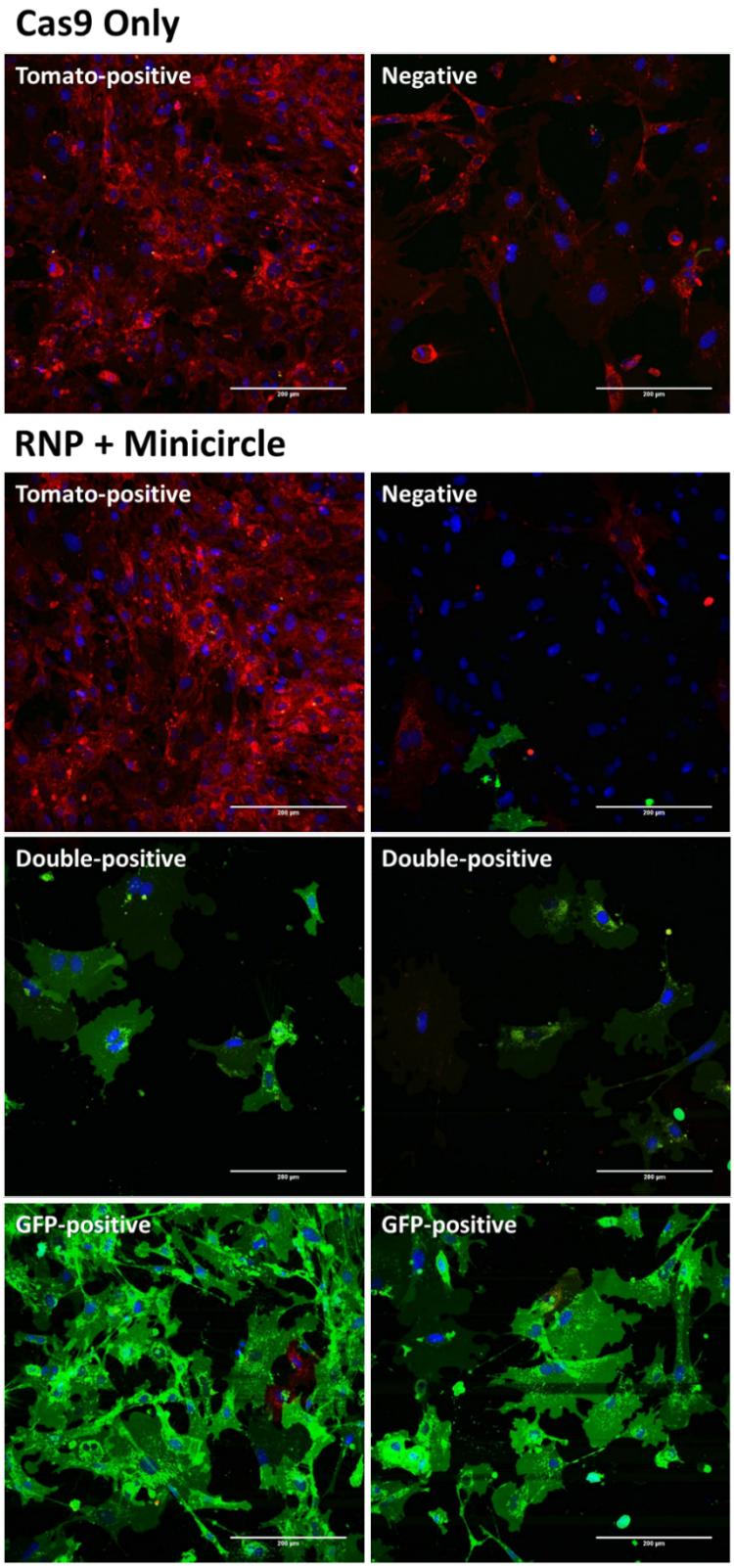


Figure 4.13: Imaging reveals a small amount of cross-over between populations of edited *mTmG* MEFs during FACS. 20x confocal images taken of each well 48 hours post-FACS. Scale bar 200µm. NB. There is some banding due to the intensity of the signal from nuclear-GFP, and to maintain laser power between wells.

4.4.2 PCR for Sequencing

Using the DNA extractions from the previous section, I performed PCR amplifications as before (Section 4.2.3). Following amplification, each PCR product was purified and 5µl was electrophoresed on a 2% agarose gel (Figure 4.14). Certain bands, e.g. D3 and E2, were very weak, although this may simply reflect that editing has removed the primer binding sites in these populations. DNA templates were coded as such:

A	-	NF-H_2O
B	-	Cas9 Only (Tomato-positive)
C	-	Tomato-positive
D	-	GFP-positive
E	-	Double-positive
F	-	Double-negative

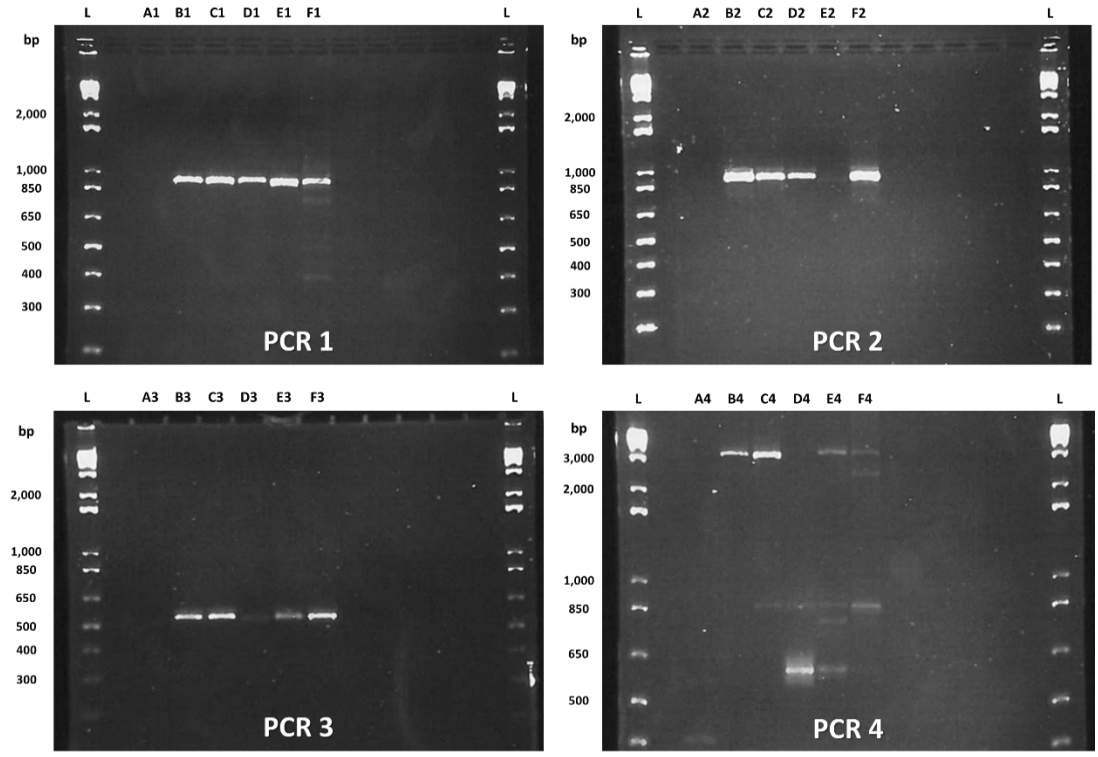


Figure 4.14: Purified PCR products for sequencing were electrophoresed on a 2% agarose gel. L = 1kb Plus DNA Ladder (Thermo Fisher).

4.4.3 Sequencing Design and Run Report

As before, purified PCR amplifications were transferred to the WTCRF for preparation and sequencing (Section 2.18). For MinION, amplicons were end repaired and adenylated, and then 16 barcode adapters were added during 18 cycles of PCR. For Ion Torrent, PCR products were sheared to a mean length of 300bp and then barcoded adapter sequences (16 in total) were added in a ligation step. Barcodes allow separation of reads by population and PCR product *in silico*. A run report was automatically generated by the Ion Torrent software upon completion of sequencing. Again, this indicated that there was a substantial decrease in quality towards the ends of the reads, and that the mean read length was significantly below the input (189 vs. 300, Figure 4.15).

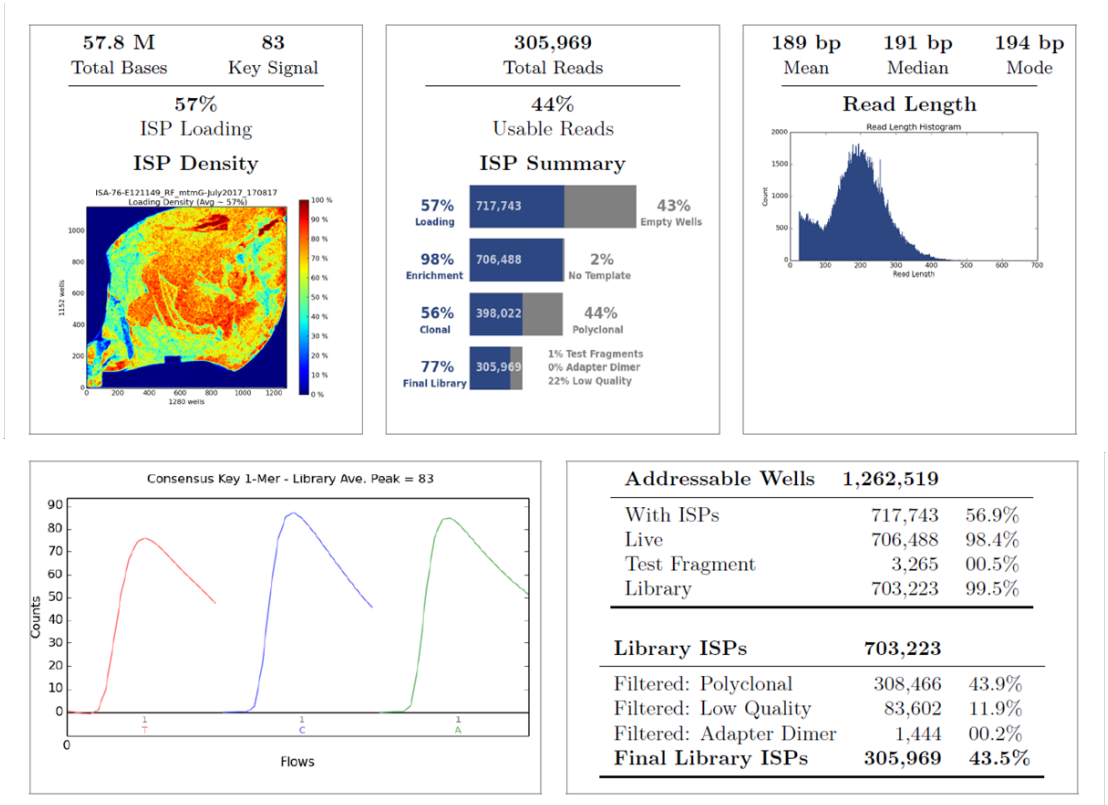


Figure 4.15: The Ion Torrent run report indicates there has been substantial trimming of the reads due to their low quality.

Following sequencing, Ion Torrent reads were returned in files separated by barcode and aligned and analysed using the pipeline described in Section 2.19.1. The MinION reads were returned unsorted, and so initially were processed with the tool Porechop

(https://github.com/rrwick/Porechop) to sort them by barcode and then remove the adapter sequences. A modified version of the Ion Torrent pipeline was then used to align and analyse the reads (Section 2.19.2). Furthermore, a *de novo* genome assembly was attempted with both data sets (Section 2.20).

4.5 Ion Torrent Sequencing Results

4.5.1 The *mTmG* System Potentially Underestimates Rates of NHEJ – Deep Sequencing Reveals Greater Than 20% of the Tomato-only Population Is Edited

I aligned the relevant barcoded reads against the reference sequences for PCR products 1 and 2, which contain the first and second loxP sites respectively. A high proportion of the reads were found to align in all cases (>90%) indicating the remaining reads were high quality and had been sorted by barcode correctly. On viewing the alignments in the IGV browser, I observed that a large proportion of the reads from the Tomato-positive sub-population contained insertions and deletions across the loxP sites. I therefore quantified the percentage of reads containing indels across the loxP site ± 10bp in each PCR product. The results showed that over 20% of the Tomato-positive reads contained indels or SNVs across the first loxP site (Figure 4.16), while ~8% of the reads were edited at the second loxP site (Figure 4.17). Unfortunately, as individual Ion Torrent reads do not span both loxP sites, it's impossible to combine these two results. However, this still indicates that a significant proportion of the visually unedited cells have genomic evidence of editing. This could arise when there is editing at only one site, or where out-of-sync editing occurs at both sites such that the intervening fragment is not excised. Indeed, there is a large downward shift in Tomato intensity (Figure 4.12B) on editing, as opposed to the more tightly clustered population in controls. As the Tomato-positive cells comprised ~75% of the total following RNP transfection (Figure 4.12B), this would suggest that the rate of NHEJ was 15% higher than detected by FACS. Nevertheless, this underestimation was expected and is not a serious concern for a model which will mainly be used to optimise reagents for *in vivo* delivery and to increase HDR.

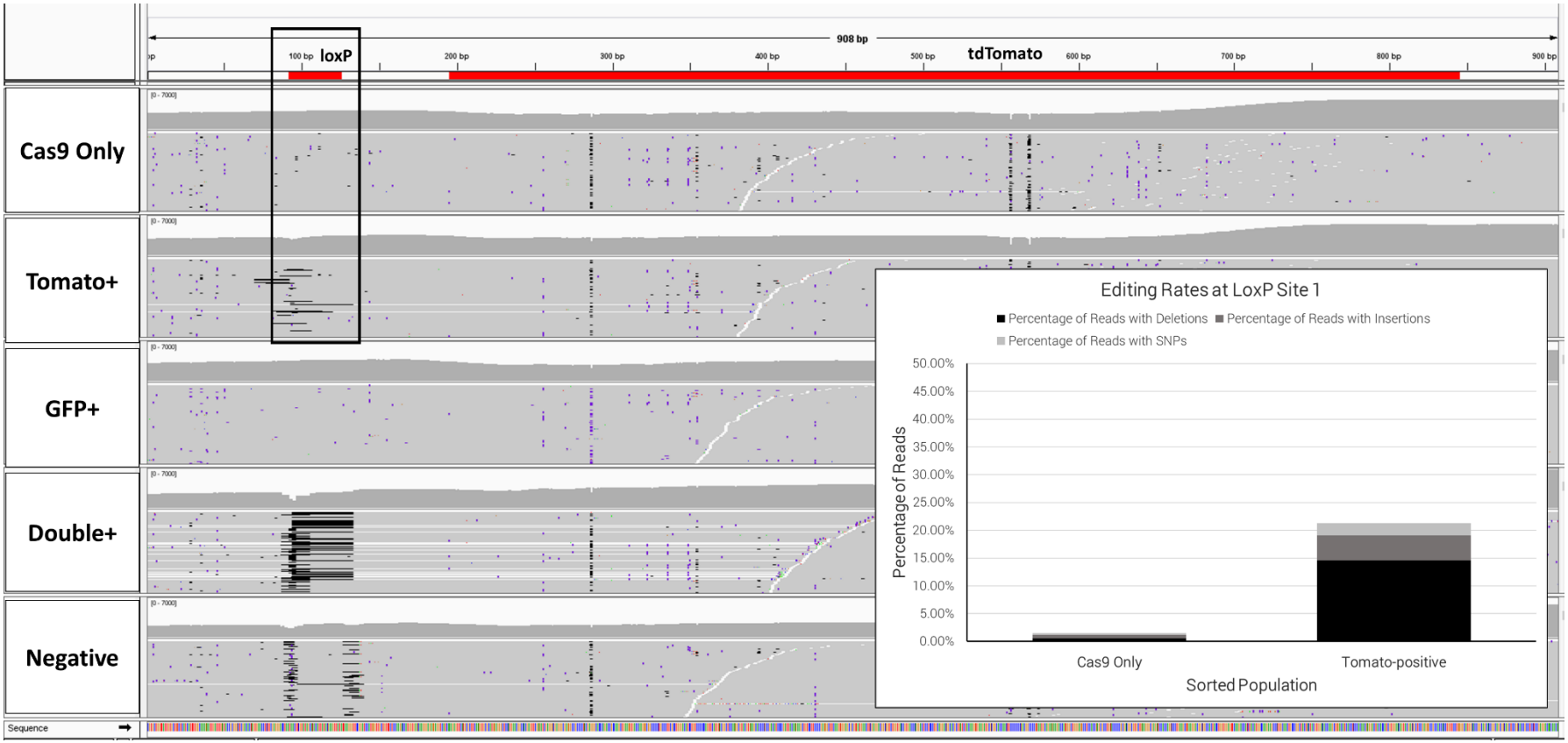


Figure 4.16: The alignments for PCR product 1 reveal a large proportion of the Tomato-positive population harbours small indels around the first loxP site. Screen capture from the IGV browser software. Regions of interest are highlighted in red. The black box indicates the region used to determine the rates of editing, which have been plotted graphically in the insert. NB. As we expect the binding site for primer 5 to be removed in the GFP-positive population it is likely the reads here are derived from the Tomato-positive contamination.

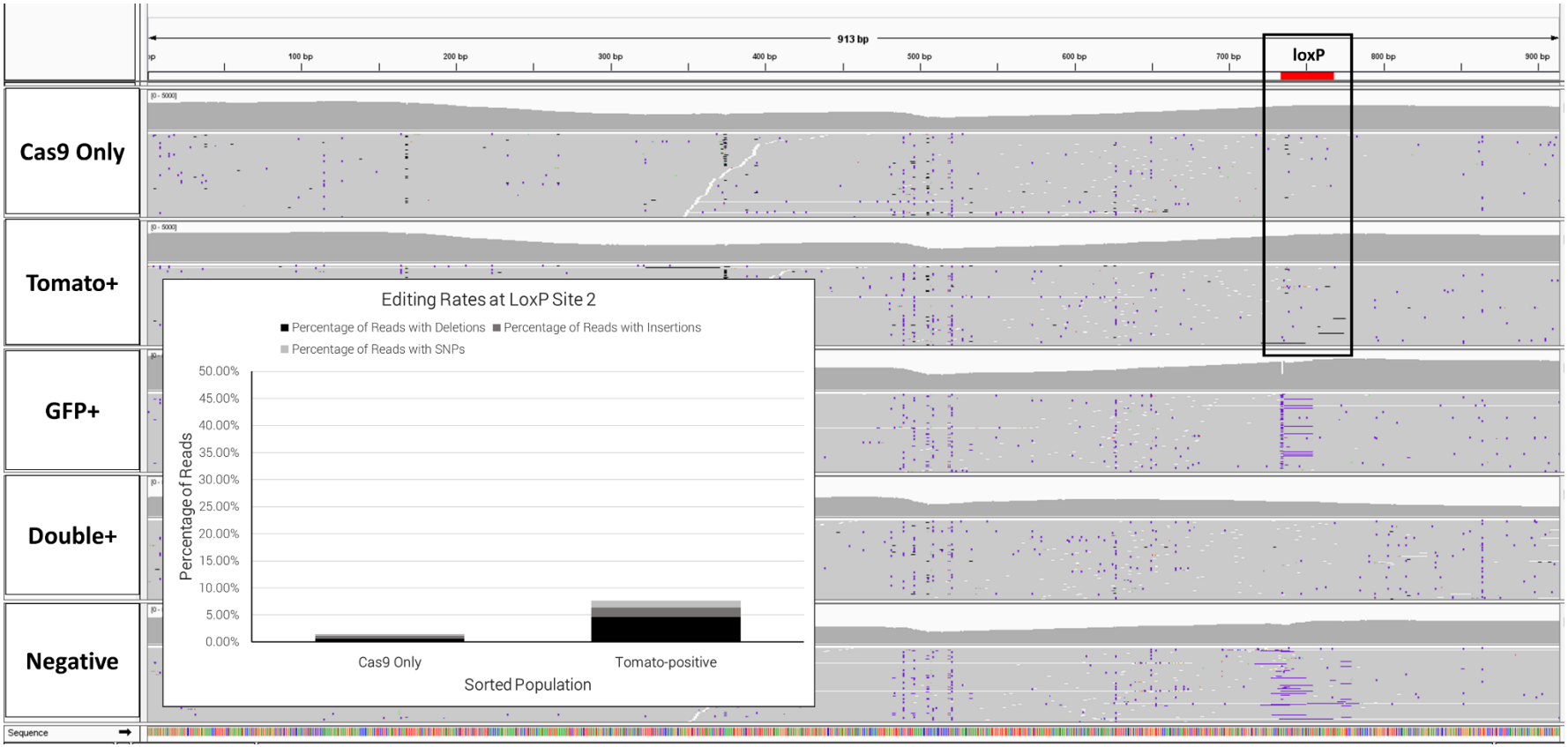


Figure 4.17: The alignments for PCR product 2 reveal a smaller proportion of the Tomato-positive population harbours indels around the second loxP site. Screen capture from the IGV browser software. The loxP site is highlighted in red. The black box indicates the region used to determine the rates of editing, which have been plotted graphically in the insert. NB. As we expect the binding site for primer 2 to be removed in the GFP-positive population it is likely the reads here are derived from the Tomato-positive contamination.

4.5.2 The Tomato-containing Fragment Between the LoxP Sites Has Been Excised in the GFP-Positive Population

Next, I aligned the reads for PCR product 4 against the expected reference sequence and the predicted sequence of the cleavage product. These support two opposite hypotheses; in the first there has been no excision of the region between the loxP sites, whereas in the second this excision is total. On viewing the alignments against PCR product 4 in the IGV browser, almost no reads from the GFP-positive population were observed to contain the sequence between the two loxP sites (maximum coverage 11x vs. 1496x for Tomato-positive, Figure 4.18). These reads may be derived from the Tomato-positive contamination of the GFP sort. Furthermore, when considering the alternative hypothesis, the generation of the cleavage product, an alignment rate of approximately 85% is seen with the GFP-positive population (Figure 4.19, insert). These results are a clear demonstration that the GFP-fluorescence read-out is indicative of the expected cleavage and end-joining in our system. The range of 'indels' which can be seen around the loxP site in the cleavage product are likely due to the different lengths of sequence which are left following cleavage by LoxP#1 or LoxP#2, along with varying amounts of end-resection. Indeed, the size of deletions which can be created by CRISPR/Cas9 is known to be highly variable (Kosicki, Rajan et al. 2017).

This analysis provides no definitive answer into why the negative population has lost fluorescence, which is investigated further in Section 4.7. However, the analysis does suggest that the double-positive population is a hybrid of events in between the Tomato-positive population and the GFP-positive population. Approximately 50% of the reads from the rare double-positive population (1%, or ~3,000 cells) align uniquely to the expected cleavage product. While less common in cultured primary cells, this could occur if the locus had been duplicated and only a single copy had been modified. Alternatively, editing at the second loxP site has occurred but has deleted back into the C-terminal end of the Tomato-cassette containing the polyadenylation sequence – this could result in a hybrid Tomato and GFP translation product. However, further experiments, such as fluorescence *in situ* hybridisation (FISH), would need to be carried out to examine these possibilities.

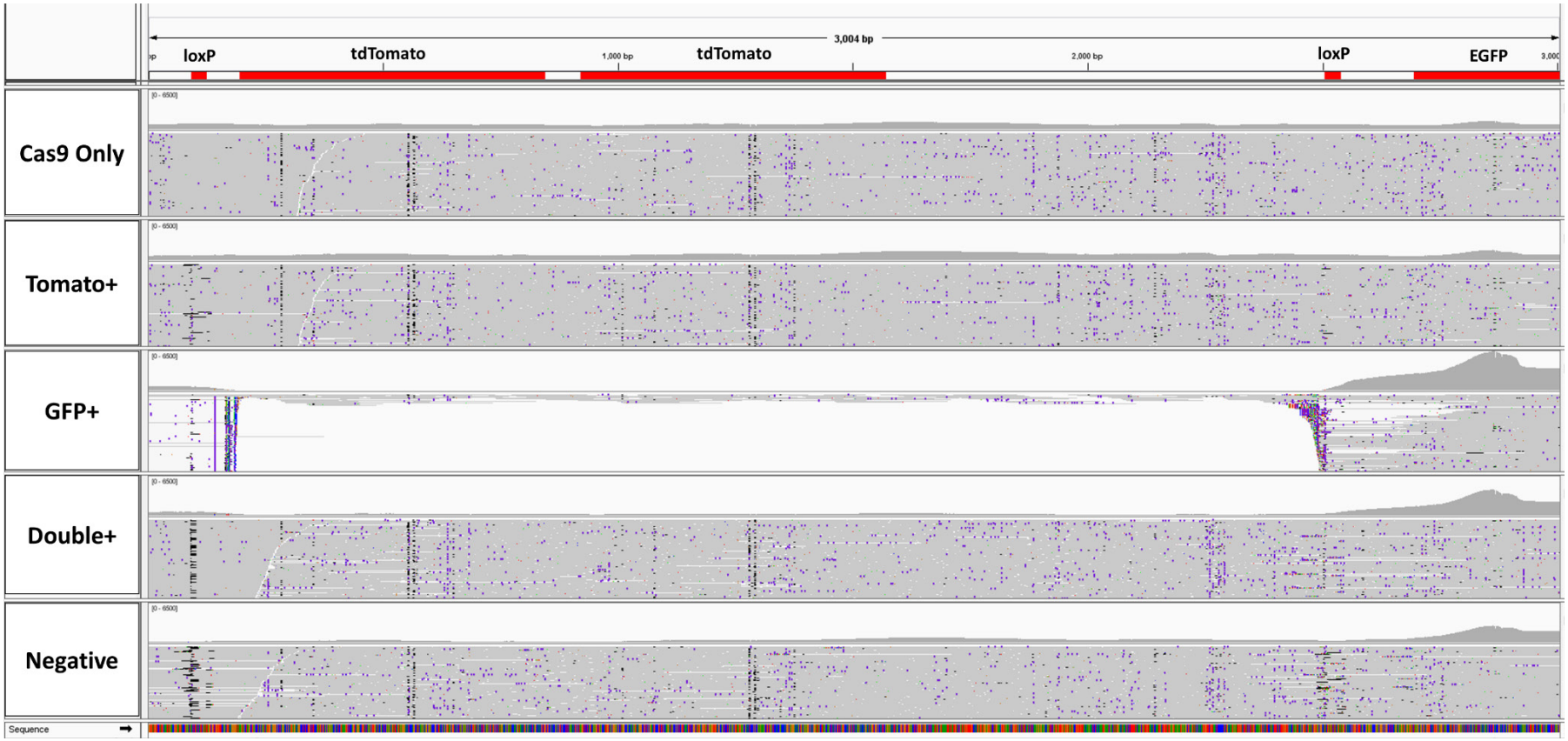


Figure 4.18: The sequences from PCR product 4 reveal that the region between the loxP sites is almost entirely absent in the GFP-positive population. Screen capture from the IGV browser software. Alignments are displayed with coverage plots above. Regions of interest are highlighted in red.

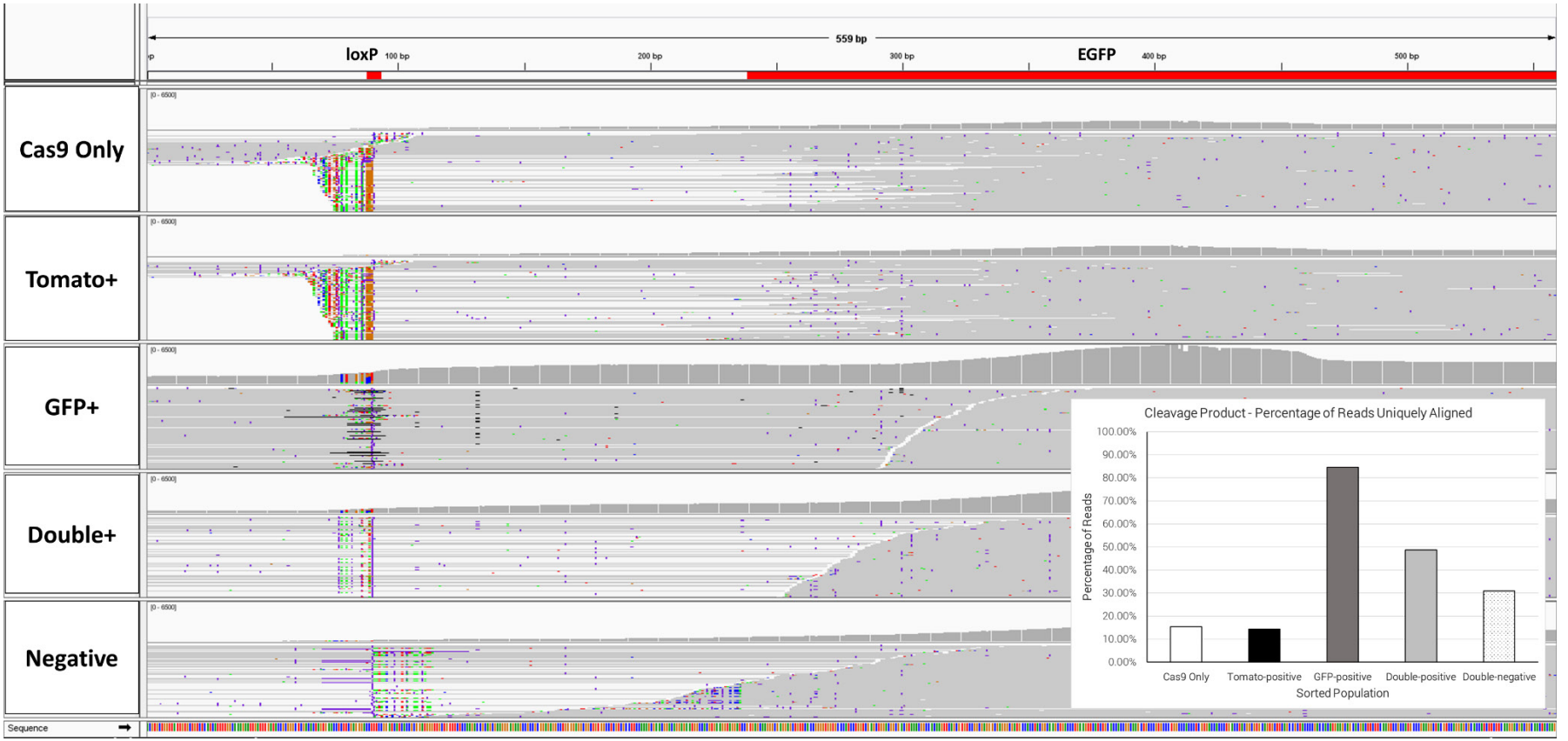


Figure 4.19: The majority of reads from the GFP-positive population align correctly to the expected cleavage product. Screen capture from the IGV browser software. Regions of interest are highlighted in red. Insert: The percentage of reads from PCR product 4 which align uniquely to the cleavage product has been plotted graphically.

4.5.3 H2B Sequence can be Detected in All Transfected Samples

Next, I aligned the reads from PCR product 4 against the predicted sequence for the H2B repair product. As expected, no reads from the Cas9 only control contained the H2B insert sequence (Figure 4.20, 4.21). This region, which contained no reads in the Cas9 only control, was then used for comparison (highlighted with a black box in Figure 21). I calculated the average read depth across this region and compared it with the total number of reads from PCR product 4 in each sample to derive the percentage of reads which include this H2B sequence (Figure 4.20). This analysis revealed that the percentage of the GFP-positive reads which contain H2B was approximately as expected, at 1.5%. However, unexpectedly, a greater number of reads from the other sub-populations contained the H2B sequence than did the GFP-positive sub-population.

Returning to the PCR products in Figure 4.14, it becomes clear that all the sub-populations contain a band at ~850bp and its intensity is proportional to the results seen in Figure 4.20. It seems likely that the seven-day period post-transfection was not sufficient to ensure the minicircle was adequately degraded and/or diluted out, allowing PCR to prime off the repair template. Alternatively, one of the reagents used for PCR may have been contaminated with the repair template. Returning to the initial Ion Torrent experiment (Figure 4.4), bands at ~850bp can also be seen in most populations, including the Cas9 only control, although strangely not in the H2Bv2 only control. In this experiment the repair plasmid would have had to persist for 17 days in a fast dividing cell population, which seems unlikely, although it is also possible that the repair templates are integrating elsewhere in the genome. However, evidence of HDR can be seen in the GFP-positive population due to the 2bp deletion which is present in the locus but not in the repair template. Very few reads from the Tomato-positive population contain this deletion, suggesting they are from the minicircle, whereas most reads from the GFP-positive cells do, indicating repair (Figure 4.22, red box).

In future experiments, in order to ensure we only detect instances of H2B which are correctly integrated into the *mTmG* locus, it will be necessary to use primers which lie outside the regions of homology in the repair template. A downside of this approach

would be the large (up to 5kb) PCR products, which may be difficult to amplify from this locus, although this could be overcome with optimisation.

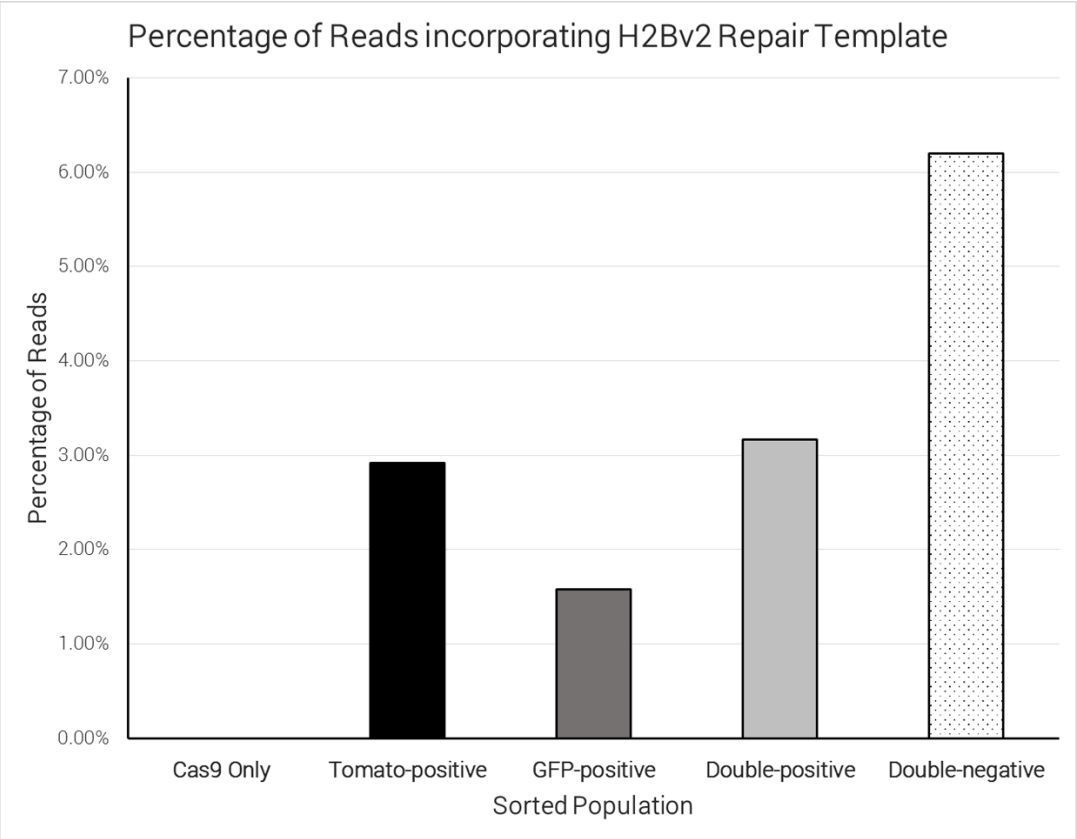


Figure 4.20: The H2B sequence can be detected in all the samples transfected with the minicircle repair template. The percentages above are derived from the average read depth across the centre of H2B, indicated by a black box in Figure 21.

Figure 4.21: In all edited sub-populations, reads can be seen to align to the H2B repair sequence. Screen capture from the IGV browser software. Regions of interest are highlighted in red. The black box indicates the approximate region used to calculate the percentages in Figure 20. The red box indicates the 2bp deletion which is present in our locus but not in the minicircle repair template.

4.5.4 *De novo* Genome Assembly is Unable to Reconstruct the *mTmG* Locus from Ion Torrent Reads

I also attempted to carry out a *de novo* genome assembly with the Ion Torrent data using the SPAdes software (Bankevich, Nurk et al. 2012). *De novo* genome assembly allows me to take a hypothesis-free approach, and let the algorithm assemble the reads through their overlapping regions into the DNA sequences from which they originate. Unfortunately, with our ~190bp reads, it was unable to produce any sequences which resembled the reference *mTmG* locus. Given the repetitive nature of the *mTmG* sequence, it would be difficult or impossible to assemble short reads unambiguously; alone, the tdTomato gene alone contains a tandem repeat which is much larger than our mean read length (~700bp vs ~200bp).

4.6 MinION Sequencing Results

The software used to align the MinION reads, Graphmap (Sovic, Sikic et al. 2016), offers three different approaches for aligning reads; global, local and semi-global or 'glocal'. In a global alignment, an attempt is made to align all the bases in the sequence and the reference, whereas in a local alignment, regions can be trimmed from the 5' and 3' ends to form a substring prior to performing the alignment. A global alignment therefore performs best when the input sequences are of a similar length to the reference sequence and have the same 5' and 3' ends. A local alignment, however, may substantially trim back reads or the reference to find the best matching portion. The hybrid method, glocal alignment, allows trimming of the reference sequence but not of the reads. It is particularly useful when one expects insertions, deletions and/or rearrangements in the reads relative to the reference. Based on this, I decided to align the reads using a 'glocal' algorithm, however, in practice, all three algorithms were found to produce similar results when tested (data not shown).

4.6.1 MinION Sequences Carry a High Error Rate

As mentioned previously, the MinION data was received unsorted by barcode, and so I initially processed it with Porechop to separate the reads by barcode and remove adapter sequences. Unexpectedly, after running this programme, the majority of reads were found to contain no identifiable barcode sequence, although there was a generally

even amount of data from the sequences with barcodes (Figure 4.22A). I analysed the sequences with no barcode further, which revealed that they had markedly reduced quality scores compared to the sequences with barcodes (Figure 4.22B). I hypothesised that this may be because these are much longer reads, however, on plotting the read distribution within the no barcode population this was not observed (Figure 4.22C). Therefore, it appears as if a large proportion of the reads from the MinION are simply too low quality for the barcodes to be identified correctly (a minimum of 75% identity was required to assign a read to a barcode).

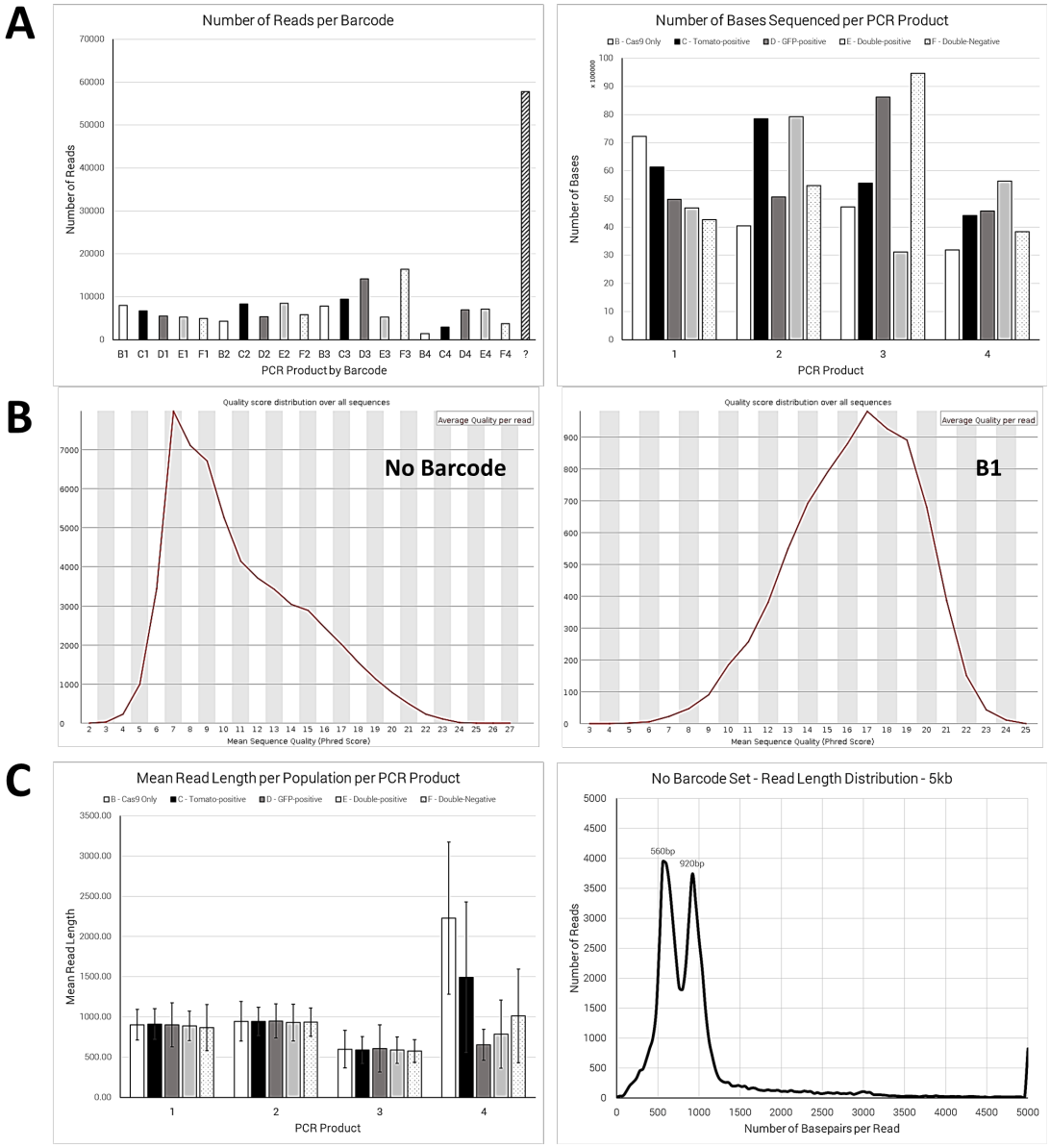


Figure 4.22: The high error rate in MinION data prevents the identification of barcode sequences. A) The amount of data generated per barcode was largely the same, although most reads had no identifiable barcode. B) Quality score plots generated by FastQC for the no barcode population and B1. C) Comparing the mean read length of each PCR product with the read length distribution of the no barcode population reveals it is mostly comprised of the smaller PCR products.

4.6.2 The MinION Sequencing Protocol is Biased towards Sequencing Shorter Products

After sorting by barcode, I examined the mean read length (Figure 4.22C) and read length distributions of each PCR product. All of the PCR products from PCRs 1 through 3 were of the expected size and had a single peak in their read length distributions. In PCR product 4, which is ~3kb as opposed to 900bp for 1 and 2, there were a range of peaks in the read length distribution, as expected based on the gel picture. However, there was a strong bias in the read length distributions towards the shorter PCR products. In several cases, the shorter PCR products were of a very low abundance on the gel and yet comprised the majority of the sequences returned (Figure 4.23). This indicates that, given a population of mixed size, the MinION will preferentially sequence the shorter products present. Based on this, it suggests that currently use the MinION to quantify editing events where one expects products to be of varying sizes is not advisable.

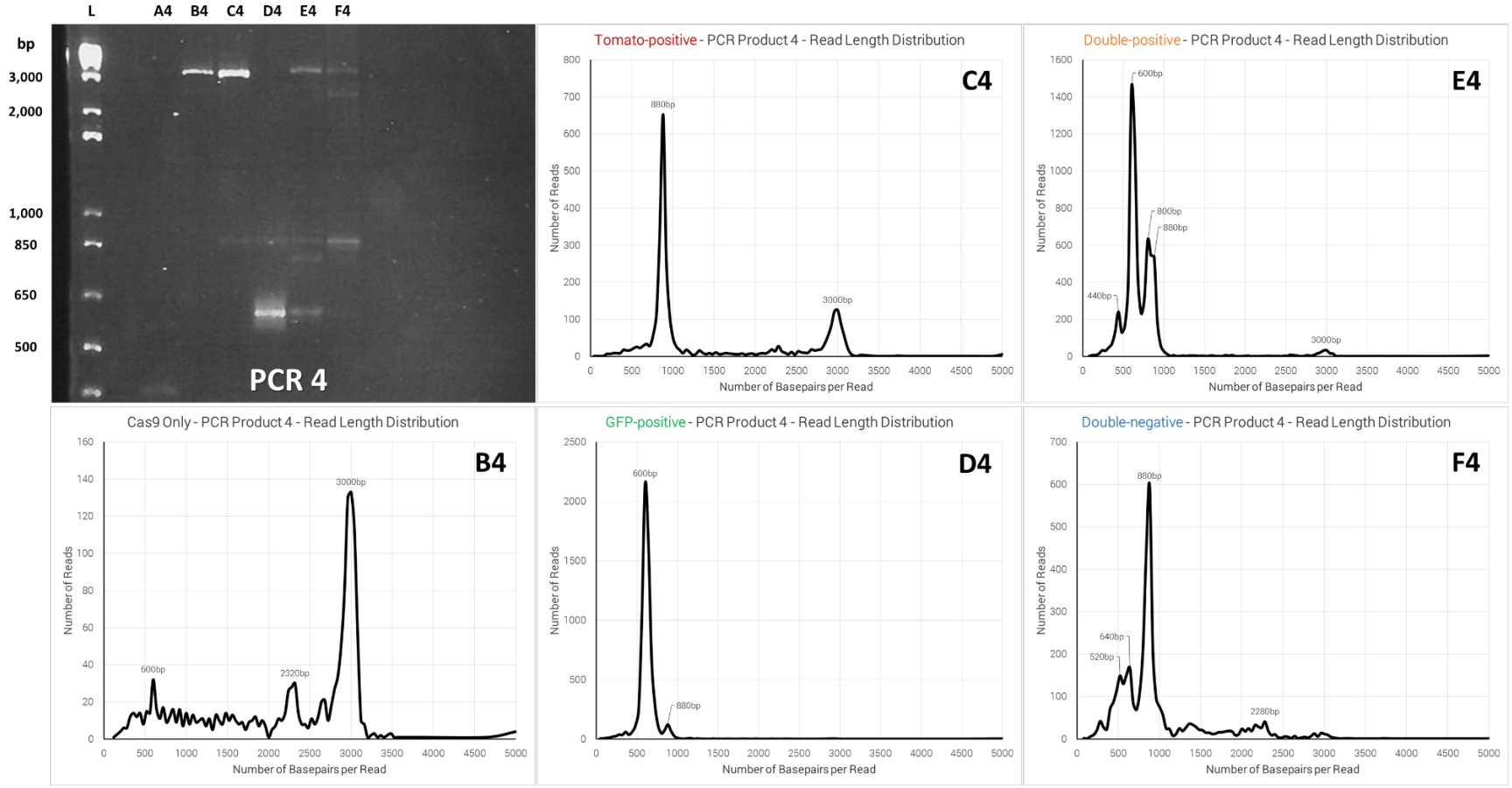


Figure 4.23: When given a population of DNA products of mixed sizes, the MinION exhibits a strong preference towards sequencing the shorter fragments. Top left: Products from PCR 4 were visualised on a 2% agarose gel, revealing bands with a range of sizes. Comparing the read length distributions with the gel picture, there is a significant discrepancy between band intensities and the frequency of sequenced products seen.

4.6.3 MinION Sequencing Confirms the Results from Ion Torrent

Next, I aligned the processed MinION sequences, with their barcodes and adapters removed, against the expected/predicted sequences as before. The alignments largely confirmed what was seen in the Ion Torrent data; in that the region between the loxP sites was excised in the GFP-positive population (Figure 4.24) to form the expected cleavage product (Figure 4.25), with the double-positive population appearing to be a hybrid between the GFP-positive and Tomato-positive populations. However, compared to the Ion Torrent data, the high error rate of the MinION was clearly evident in the alignments. These errors were evenly distributed across the reads and make it impossible to discern whether there has been editing at the loxP sites, as was done with the Ion Torrent data.

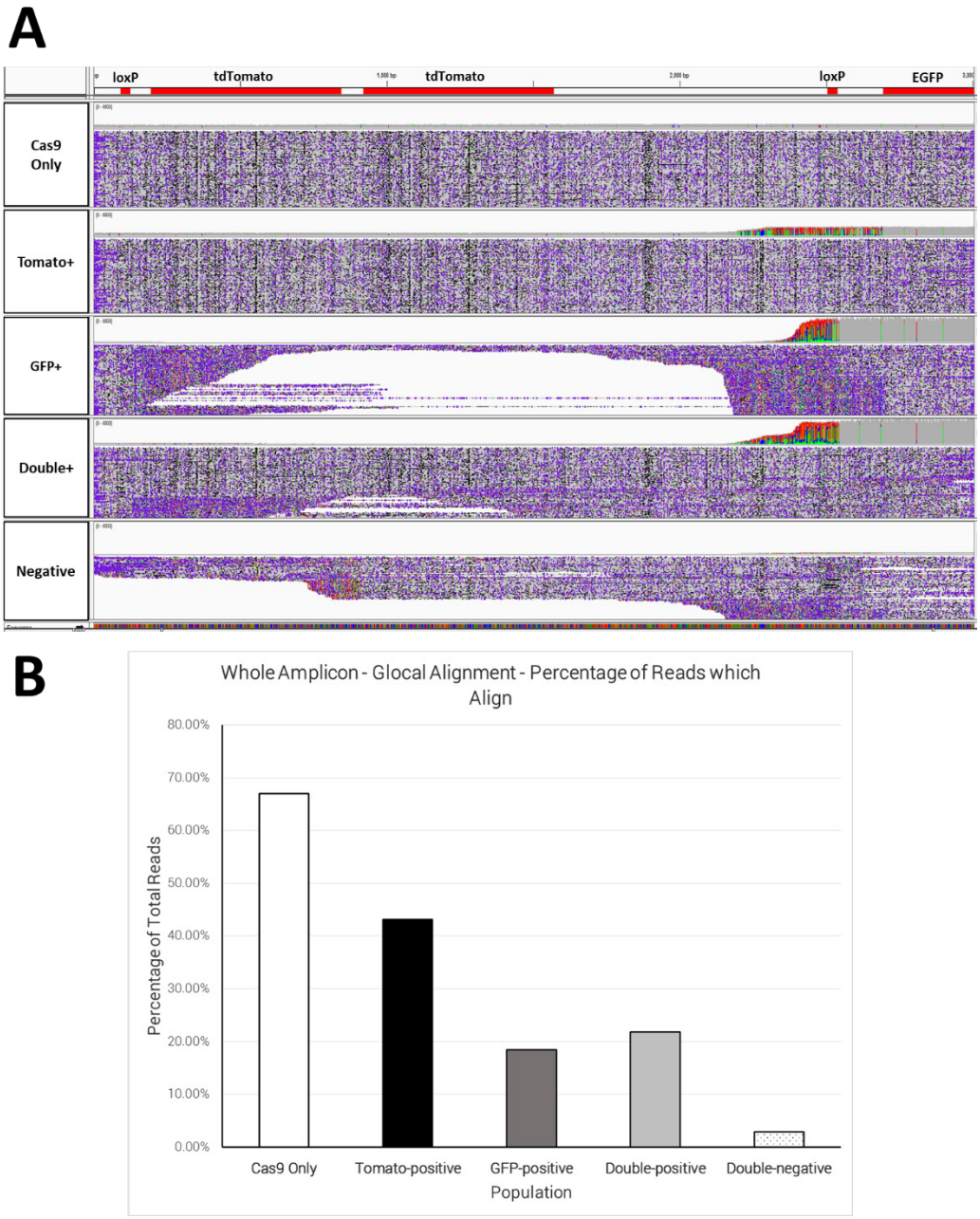


Figure 4.24: The MinION data confirms that the region between the loxP sites in the GFP-positive population is removed. A) Screen capture from the IGV browser showing the alignments and coverage plots for each population. Regions of interest are highlighted above in red. B) The percentage of reads from each population which align to the reference for PCR product 4.

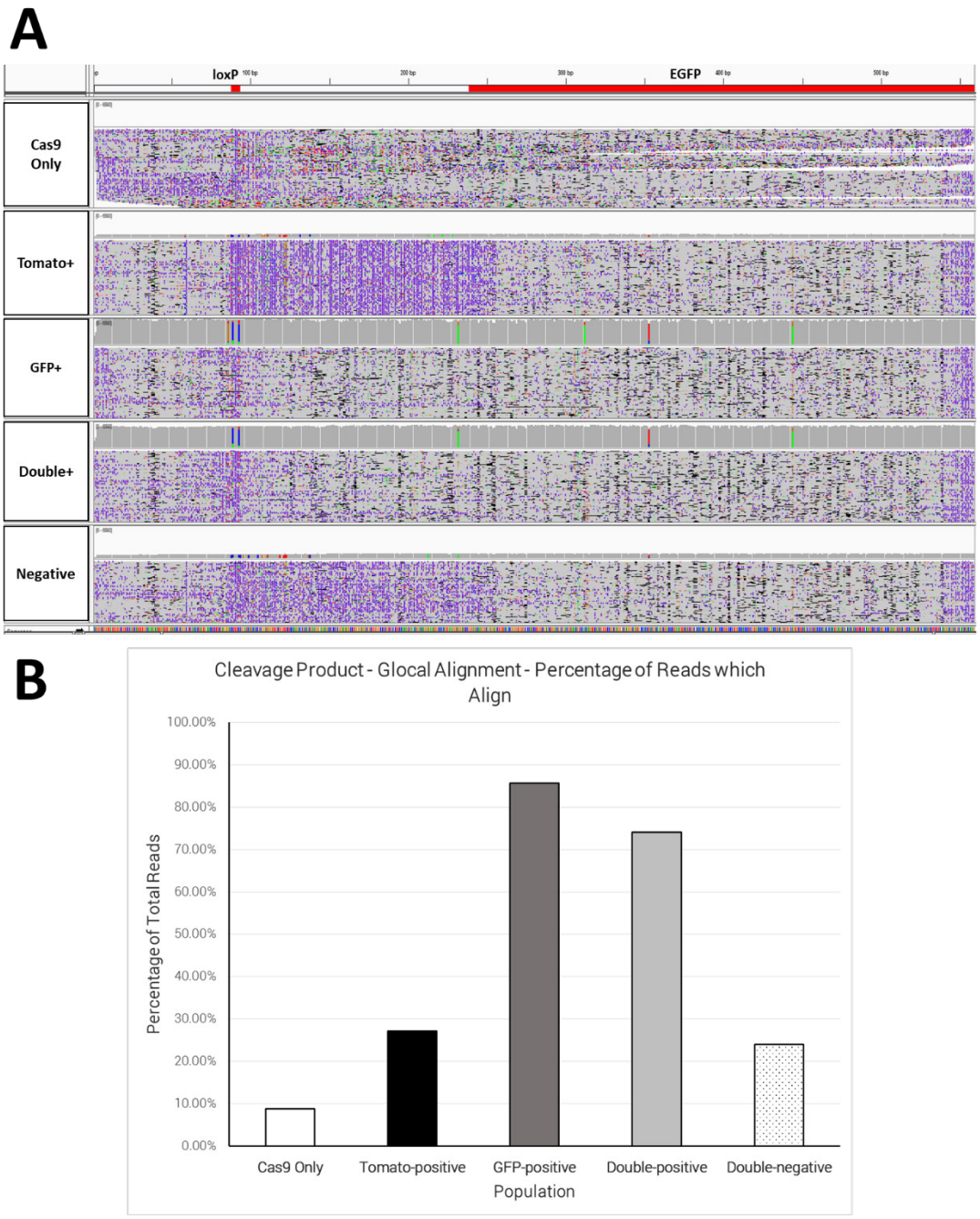


Figure 4.25: The expected cleavage product is present in the GFP-positive population, and the double-positive population. A) Screen capture from the IGV browser showing the alignments and coverage plots for each population. Regions of interest are highlighted above in red. B) The percentage of reads from each population which align to the expected cleavage product.

4.6.4 *De Novo* Genome Assembly Is Successful with Longer MinION Reads

As the MinION sequencing platform produces much longer reads than the Ion Torrent, it should be highly suited to performing *de novo* genome assemblies. MinION reads should extend beyond any repetitive regions, allowing them to be correctly ordered and assembled. Considering this, I attempted a *de novo* genome assembly with the reads from PCR product 4 for each of the samples using an algorithm, Canu (Koren, Walenz et al. 2017), which is optimised for MinION data.

The results revealed that the MinION reads can be successfully assembled in order to form the major sequences which are present in the input. Aligning the resulting sequences on SnapGene to the reference, it could be seen that Canu had assembled the reference sequence from the Cas9 only control with very little error (Figure 4.26A). Furthermore, the three sequences assembled from the GFP-positive data were all lacking the tdTomato. Interestingly, Canu identified that the double-positive population consists of a mixture of sequences with Tomato (88%) and without (12%), while the double-negative population was also seen to be a mixture, with some sequences missing segments of the Tomato reporter; which could explain their loss of fluorescence.

From the SnapGene alignments, I noticed that a large proportion of the edited sequences contained an insertion just upstream of the GFP in place of the MARCKS tag (noted by a small blue arrowhead). I hypothesised that this might be indicative of the H2B sequence being present, and so the assembled sequences were aligned again, this time to the expected repair product. This analysis revealed that the insertion was indeed H2B, although again it was detected in all the samples transfected with the minicircle repair template (Figure 4.26B), indicating possible sequencing off the repair template as opposed to the locus.

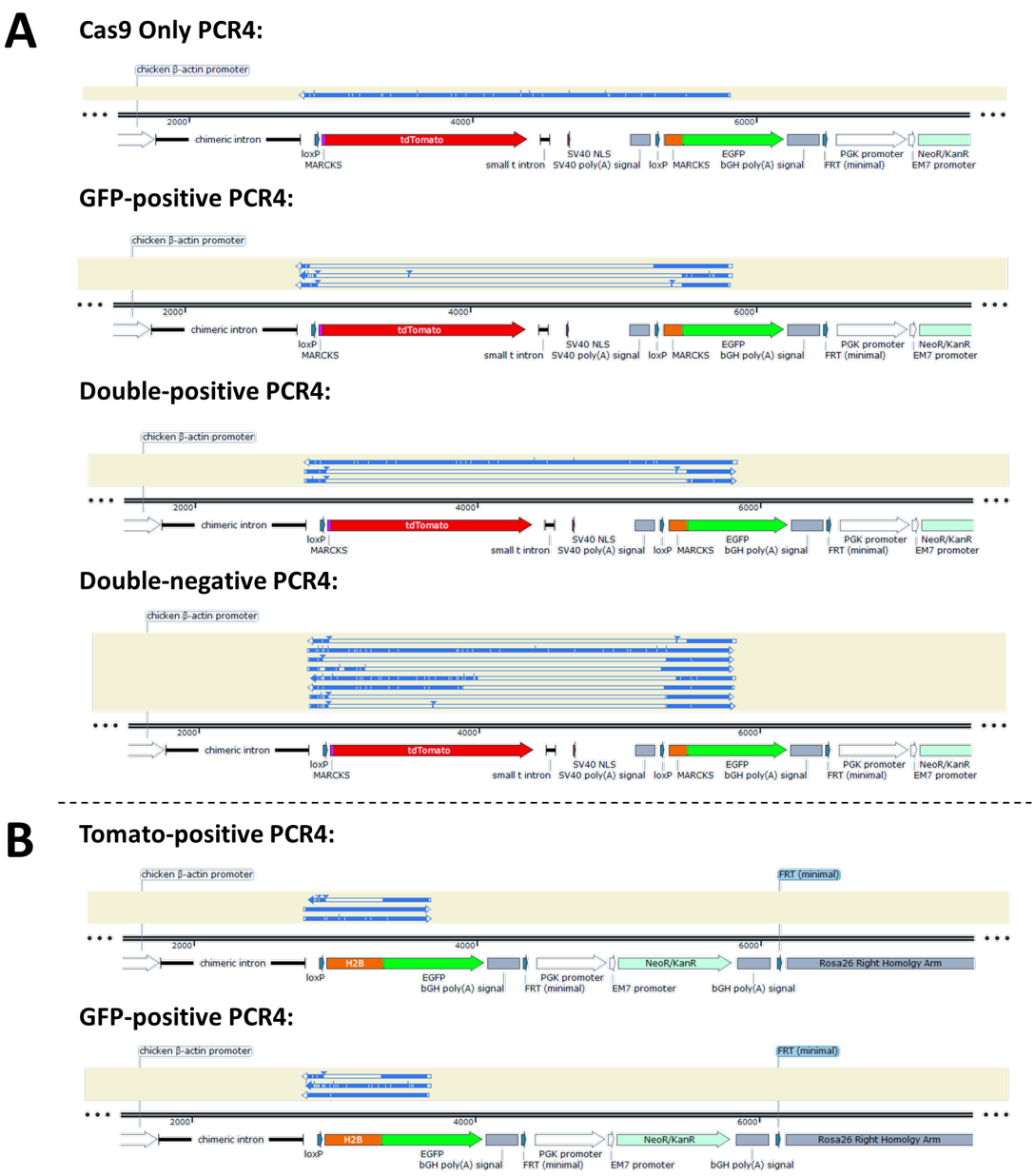


Figure 4.26: A *de novo* genome assembly was performed on the MinION data for PCR 4 using the Canu software. The resulting assemblies were aligned against the expected sequences using SnapGene, and alignments are shown above for PCR 4 (A) and the H2B repair sequence (B). Assembled sequences are ordered based on the number of reads which went to form them, with the assembled sequence which represents the largest number of reads uppermost.

4.6.5 H2B Sequence Is Again Detected in All Minicircle Transfected Samples

Following on from the detection of H2B in all samples by *de novo* genome assembly, alignments of the reads from PCR 4 against the expected repair product produced a

similar result. H2B was detected in a large and reasonably equal percentage of all the populations transfected with the minicircle (Figure 4.27).

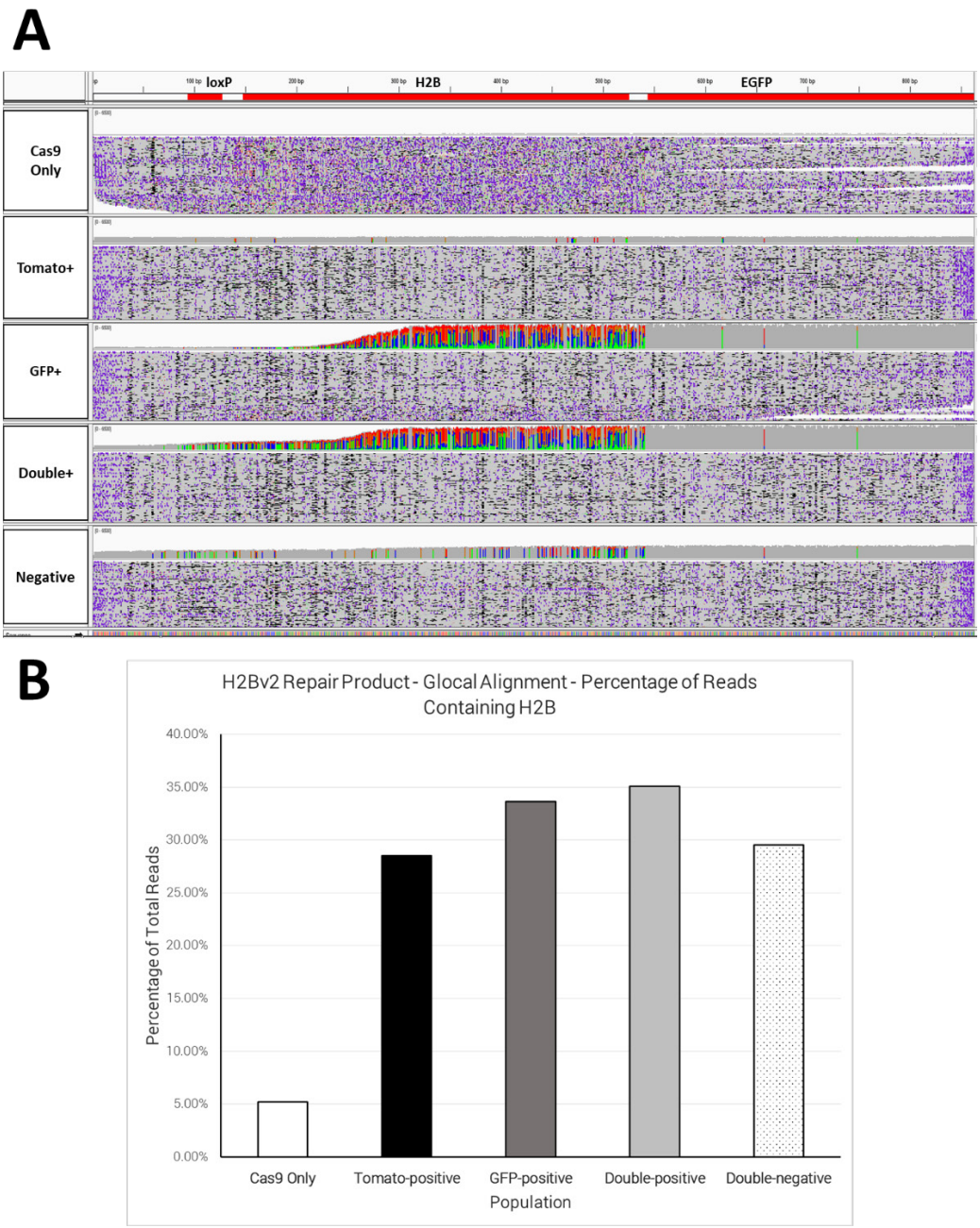


Figure 4.27: The H2B repair sequence is detected in all samples transfected with the minicircle repair template. A) Screen capture from the IGV browser showing the alignments and coverage plots for each population. Regions of interest are highlighted above in red. B) The percentage of reads from each population which contain the H2B sequence. Percentages calculated as described in Section 4.5.5.

4.7 TOPO-Cloning of the Double-Negative Population Reveals Large Deletions across the Promoter and/or GFP

As next-generation sequencing had provided no definitive explanation for the negative population, it was hypothesised that this group must contain larger deletions which extend beyond the binding sites for our current sequencing primers. Based on this, I designed new sequencing primers which lie within the pCA promoter and further into the GFP (Section 2.5.2 for sequences). PCR amplifications were then performed using these primers with DNA from both the Cas9 only control and double-negative populations. Amplifications were visualised on a 2% agarose gel, which revealed the presence of multiple bands in the negative population (data not shown). Following this, each PCR product was TOPO-cloned (Section 2.8) and 96 clones were submitted for Sanger sequencing. The returned sequences revealed that almost all negative cells harbour deletions that extend into the GFP sequence and/or the promoter, which would be expected to result in a loss of fluorescence. Furthermore, all sequences from the negative population were seen to have lost the tdTomato cassette (Figure 4.28).

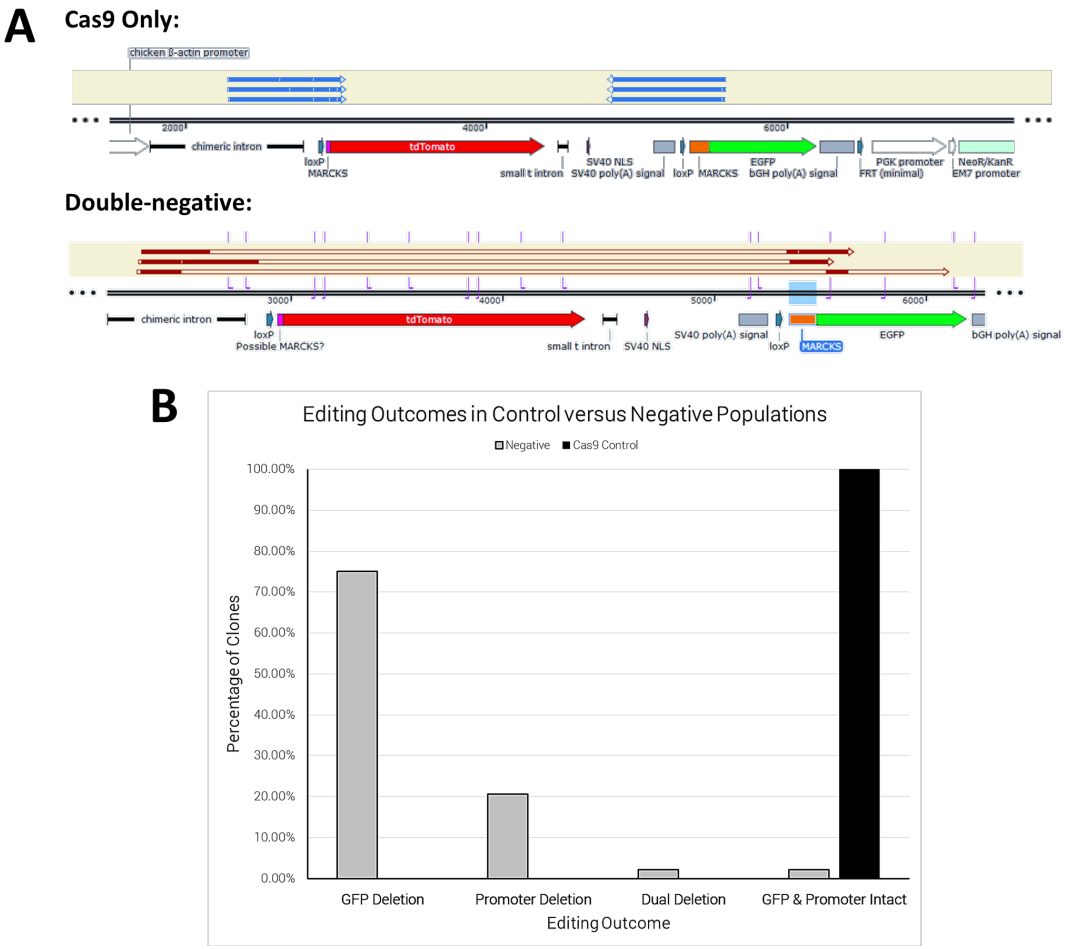


Figure 4.28: The negative population contains larger deletions which extend into the promoter and/or GFP. A) Example sequences aligned using SnapGene which show large deletions in the negative population relative to the control. B) Based on the SnapGene alignments, the number of sequences in each population which contained deletions was counted and plotted graphically.

4.8 Conclusions

With these sequencing results, it is possible to explain how the four spectrally distinct populations have arisen based on the editing they have seen at a genomic level. The GFP-positive population has undergone the expected cleavage at both loxP sites, removing the Tomato reporter, while the negative population appears to have undergone a similar process except larger deletions have been created at the loxP sites; disrupting the GFP and/or pCA promoter. It is also important to note that over 20% of the Tomato-positive cells contained indels around the cut sites, indicating that we are achieving a much higher rate of NHEJ than we can visually track. Furthermore, as a working model, I would suggest that the rare double-positive population contains more

than one copy of the *mTmG* locus, which are not equally edited. My data supports this model as the double-positive population is much reduced in non-immortalised cells which are more karyotypically stable. Nevertheless, some genome instability is expected when culturing even non-immortalised cells over large periods of time. In future work this could be tested by sequencing the mRNA transcripts from the double-positive cells; if both pCA-Tomato and pCA-EGFP transcripts are detected this would support this hypothesis. Alternatively, we may detect hybrid pCA-Tomato-EGFP fusion transcripts as a result of more complex genomic editing at the (non-duplicated) original locus.

The data presented in this chapter also allows us to draw comparisons between the two sequencing technologies employed, Ion Torrent and nanopore sequencing. Both can be seen to have their advantages and disadvantages. Strikingly, the error rate is still very high in MinION data, which can significantly hamper any further processing and analysis. However, due to the volume of data one receives from a MinION run, it is possible to filter out the lower quality reads and still be left with a large number of long, reasonably high-quality reads, which seem particularly suited to *de novo* genome assembly. Nevertheless, the error rate in the filtered reads is still higher than in Ion Torrent data, and the platform also displayed a strong preference to sequencing shorter fragments. These two factors currently mean it is ill-advisable to solely use the MinION when quantifying an editing experiment. In these cases, the Ion Torrent should give a much more representative picture of the input, as the DNA is sheared to the same size prior to being sequenced. Alternatively, TOPO-cloning is a valid approach if one expects there will be a high proportion of edited cells in the input (>1%, or ideally >10%), although it can be laborious and time-intensive if many samples are to be sequenced. It is therefore prudent to think carefully about the variables of the experiment prior to deciding upon a sequencing methodology, to ensure the experiment provides meaningful results.

Chapter 5 Correcting Pathogenic Indels in *Zmynd10* in a Mouse Model of Primary Ciliary Dyskinesia

5.1 Introduction

As detailed in the previous two chapters, I have developed a system which will allow us to develop and optimise reagents and protocols for gene editing by CRISPR in cells and *in vivo*. Given that rates of gene editing and repair have been shown to be influenced by sequence, chromatin configuration and expression level (Chen, Rinsma et al. 2016, Tsouroula, Furst et al. 2016, Jensen, Floe et al. 2017), I next had to confirm that methods which achieve high rates of NHEJ and HDR in the exogenous *mTmG* cassette in the *Rosa26* locus are also effective when targeting a disease-causing mutation in a gene of interest. Once we have developed methodologies for testing gene editing in organotypic cultures and *in vivo*, it will also be important to demonstrate an ability to edit clinically relevant cell types for our disease of interest in the reporter, at levels that would be therapeutically beneficial. As a proof-of-principle, I began to address these issues using a mouse model of the rare, highly polygenic, respiratory disease, primary ciliary dyskinesia (PCD).

Primary ciliary dyskinesia (OMIM: 244400), is a genetically heterogeneous, autosomal recessive condition which affects approximately 1 in 10,000 to 1 in 20,000 individuals world-wide (Mirra, Werner et al. 2017). PCD presents with a range of adverse symptoms, including respiratory distress, impaired mucociliary clearance, hydrocephaly, sinusitis and male infertility, due to a lack of correctly functioning motile cilia (Daniels and Noone 2015). Motile cilia are found on many of the body's epithelial surfaces, such as the airways and the brain ependyma, where they are crucial for fluid and/or mucus clearance (Horani, Ferkol et al. 2016). Furthermore, motile cilia are present on the embryonic node where they are necessary to generate the fluid flow which establishes left-right asymmetry. Because of this, approximately half of PCR patients also present with laterality defects, most commonly situs inversus, in which organs are mis-placed in their mirror-image (Damseh, Quercia et al. 2017, Mirra, Werner et al. 2017).

Cilia are complex, microtubule-based organelles which project out from the cell, to which they are linked by the basal body. In mammals, there are two varieties: a) sensory, or primary, cilia which are immotile; and b) motile cilia which are involved in generating fluid flow or propelling sperm. The cilium is a dynamic structure, and its formation and maintenance are dependent on the bi-directional trafficking of many proteins in and out of the organelle via the specialised intraflagellar transport (IFT) system. Enclosed by the ciliary membrane, the axoneme of the motile cilia consists of an array of microtubules with a ring of 9 microtubule doublets and 2 central singlets, a (9 + 2) arrangement (Mitchison and Valente 2017) . Movement is generated by outer and inner dynein arms (ODA and IDA) which periodically decorate the A-tubule of each microtubule doublet ring (Figure 5.1). ODAs control beat frequency while IDAs control waveform in an ATP-dependent dynein-powered stroke whereby the dynein "walks" along the B-tubule, allowing the axoneme to bend (Bustamante-Marin and Ostrowski 2017).

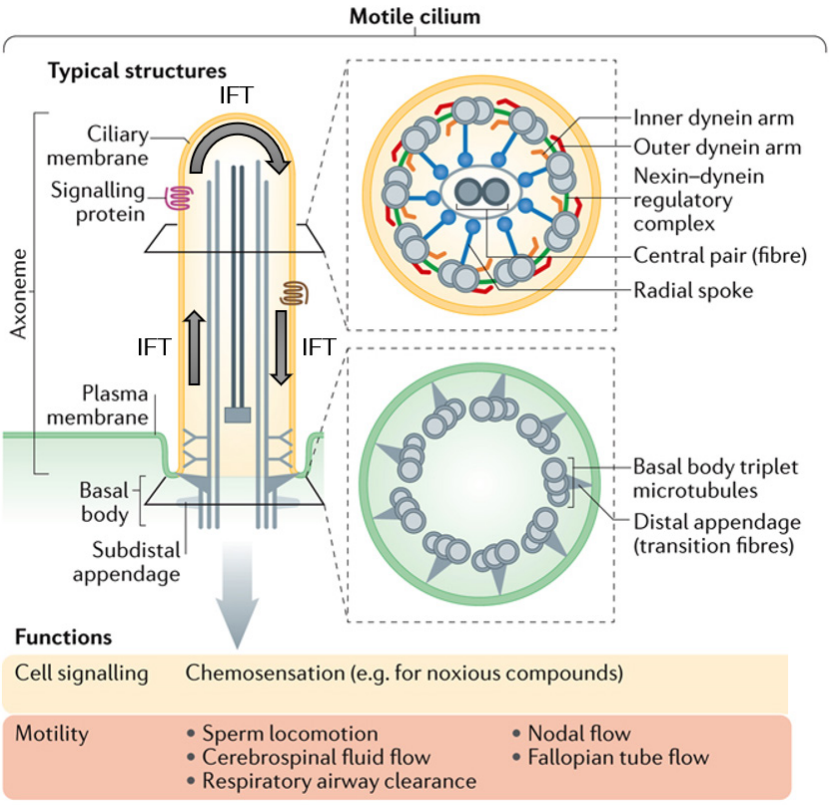


Figure 5.1: Diagrammatic representation of a motile cilium, as found on the brain ependyma and respiratory epithelium. Embryonic nodal cilia lack the central pair. Adapted from Reiter and Leroux (2017).

Immotile cilia, and hence PCD, can be the result of defects in a range of proteins, including structural elements of the cilium itself or factors involved in the assembly and transport of complexes to the cilia (Horani, Ferkol et al. 2016, Mitchison and Valente 2017). Indeed, mutations causative of PCD have currently been described in 33 separate genes (Mirra, Werner et al. 2017), a number which is continuously rising. This genetic heterogeneity, combined with the rareness of the disease, means treatment options for PCD are severely limited, and largely only symptomatic (Polineni, Davis et al. 2015). While PCD may be amenable to a gene therapy by gene transfer approach, it is a much less attractive target than other, less rare, diseases, such as cystic fibrosis (CF), which are caused by mutations in a single gene. Furthermore, the very large size of genes commonly mutated in PCD (e.g. DNAH5 at ~14.6kb) impedes packaging in the viral vectors commonly employed for gene replacement, which have a maximum capacity of ~8kb (Dunbar, High et al. 2018). Therefore, the treatment of PCD and other rare genetic diseases may particularly stand to gain from the development of CRISPR for gene therapy, as it will allow mutations virtually anywhere in the genome to be corrected with the same machinery in a "personalised" manner; potentially obviating the need for individual mutation-by-mutation clinical testing. Furthermore, if efficient, a genome editing approach allows the correction of the gene at its endogenous locus, ensuring correct regulation.

5.1.1 ZMYND10 and Ciliary Assembly

As our proof-of-principle model for PCD, we chose a *Zmynd10* mutant mouse. The disrupted gene, *Zmynd10*, codes for an essential cytoplasmic dynein assembly factor (DNAAF), required for the formation and stability of both outer- and inner dynein arms. If we can restore ZMYND10 function by genome editing to rescue axonemal dynein stability in a DNAAF model, we would likely be able to rescue all other classes of mutants. This mouse model of PCD contains a -7bp deletion in the fourth, critical exon of *Zmynd10* (*Zmynd10* c.695_701 p.Met178Ilefs*183) which creates a null allele (Figure 5.2A, B). This mouse line is part of an allelic series generated by three separate gRNAs targeting this exon in a CRISPR-mediated mutagenesis strategy which all display PCD-like symptoms, indicating that the phenotype is not due to off-target mutations. *Zmynd10* mutant mice were shown to lack ODA and IDA, and have immotile cilia (Figure 5.2) (Mali, Yeyati et al. 2017). A similar ciliopathy phenotype

in *Zmynd10* nulls has been described in several other organisms, including in human patients (Moore, Onoufriadis et al. 2013, Zariwala, Gee et al. 2013, Kurkowiak, Zietkiewicz et al. 2016), in *Drosophila melanogaster* (Moore, Onoufriadis et al. 2013), and in zebrafish (*Danio rerio*) (Zariwala, Gee et al. 2013).

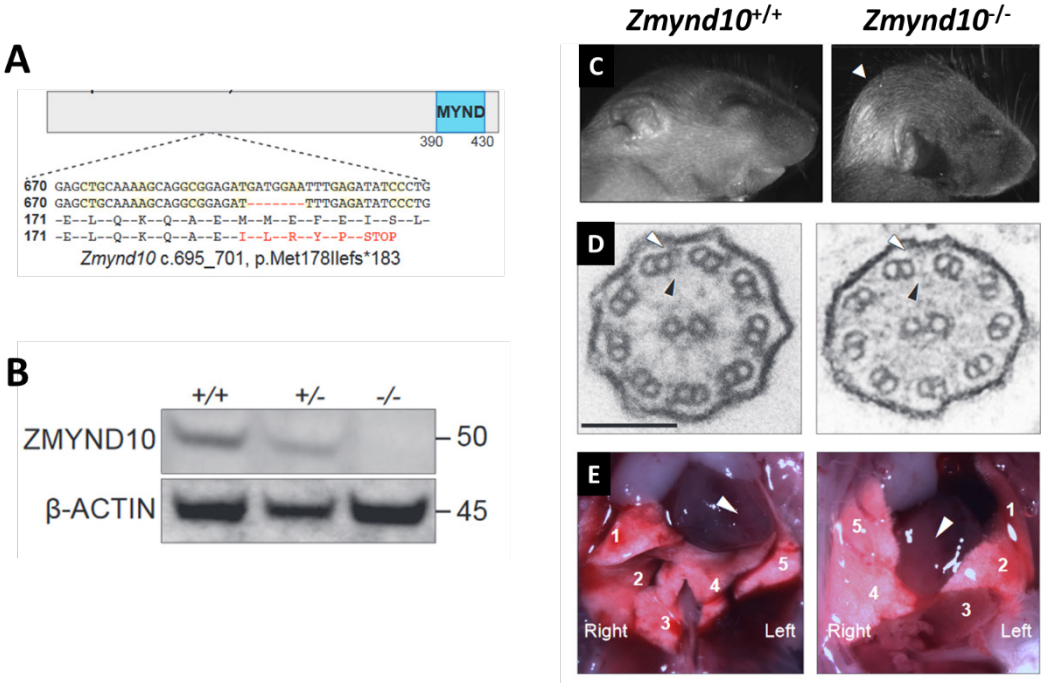


Figure 5.2: *Zmynd10* null mice display a characteristic PCD phenotype. A) Gene and protein sequence of the *Zmynd10*$^{-/-}$ mice which harbour a -7bp deletion creating a premature termination codon (PTC). B) Immunoblots from testes extracts show loss of ZMYND10. C-E) *Zmynd10* null mice display characteristic PCD phenotypes, such as C) hydrocephaly (doming of the skull, white arrowhead), D) lack of ODA (white arrowhead) and IDA (black arrowhead), and E) situs inversus (arrowhead indicates heart, numbers indicate lobes of the lung). Adapted from Mali, Yeyati et al. (2017).

The gene *ZMYND10* (GRCh38: ENSG00000004838), or *BLU*, was initially described as a tumour suppressor which was particularly prone to silencing by promoter hypermethylation in lung cancer (Marsit, Kim et al. 2005, Yau, Lung et al. 2006). Later, whole exome sequencing of families affected by PCD revealed *ZMYND10* mutations to be causative in several unexplained cases. Further immunofluorescence experiments demonstrated that the light- and heavy-chain axonemal components DNALI1 and DNAH5 were absent in ZMYND10 mutants (Moore, Onoufriadis et al. 2013, Zariwala, Gee et al. 2013, Mali, Yeyati et al. 2017). Moreover, *Zmynd10* expression was found to be controlled by RFX and FOXJ1, transcription factors known

to be required for ciliogenesis (Moore, Onoufriadis et al. 2013). These two initial studies also used overexpression of *Zmynd10* to identify a potential interaction partner, LRRC6, itself another PCD gene, with which ZMYND10 was found to co-localise in the cytoplasm, forming a putative cytoplasmic assembly complex for dyneins (Moore, Onoufriadis et al. 2013, Zariwala, Gee et al. 2013).

In order to better understand the disease mechanisms underlying *Zmynd10* mutations, our group has undertaken a variety of candidate protein-affinity purification approaches, as well as unbiased label-free quantitative mass spectrometry, to establish that ZMYND10 acts as a specificity factor for the ubiquitous chaperone protein HSP90 and co-chaperone peptidylprolyl isomerase FKBP8 for the 'client' axonemal dynein heavy chains. Interestingly, these two chaperones are critical for the folding of several proteins, such as the cystic fibrosis transmembrane receptor (CFTR) (Banasavadi-Siddegowda, Mai et al. 2011, Hutt, Roth et al. 2012), on the cytosolic face of the endoplasmic reticulum. This suggests that biosynthesis and/or folding of the large and structurally complex axonemal dynein heavy chains may be spatially restricted within the cell to regions with high density chaperones and protein quality control mechanisms by ZMYND10.

ZMYND10 appears to function in a sequential chaperone relay wherein the 'client', the heavy-chain dynein DNAH5, was found by endogenous affinity purification to initially associate with ZMYND10 and FKBP8, whereas when the 'client' was assembled into its final macromolecular complex with the dynein intermediate chain DNAI1, interactions with ZMYND10 and the chaperones were no longer detected (Mali, Yeyati et al. 2017). Moreover, once axonenal dynein arms are trafficked and docked into growing axonemes, during differentiation, the cytoplasmic levels of the dynein subunits and the assembly factors drop below the level of detection in mature cells. These findings suggest that there is a series of transient interactions, separated in time and space, during axonemal dynein assembly. As such, it will be important to test whether correcting *Zmynd10* in mature ciliated cells, the predominant cell type in the conducting airways (Bustamante-Marin and Ostrowski 2017), can restore ciliary motility. Based on these findings, I would hypothesise that successful and long-lasting functional gene correction will require targeting the basal stem cells, where cytoplasmic dynein assembly is occurring upon differentiation (Spassky and Meunier

2017). Considering this, the *mTmG* system we have developed should provide invaluable preliminary results regarding which cell types we are able to edit with various protocols.

5.2 *Zmynd10* Guide and Repair Template Design

Guide-RNAs were designed using a suite of online tools (Section 2.1.2) and chosen based on their high scores and low number of potential off-target sites. Three guides were chosen to target the -7bp deletion locus, one on the forward strand and two on the reverse strand (Figure 5.3A). All guides were initially cloned (Section 2.2) into the pX458 Cas9-GFP expression vector (Section 2.2.2), to allow for selection.

Repair templates were designed for each guide and synthesised by IDT in ssODN format. For the 'Z10 R' guide, two alternative repair template designs were tested, one with approximately symmetrical regions of homology on either side of the Cas9 cut site, and one with asymmetric regions of homology. For the other two guides only asymmetric repair templates were designed (Figure 5.3B), as asymmetric ssODN repair templates have been reported to drive higher rates of HDR (Richardson, Ray et al. 2016). All repair templates carried silent mutations across the PAM site in order to prevent re-cleavage of repaired alleles.

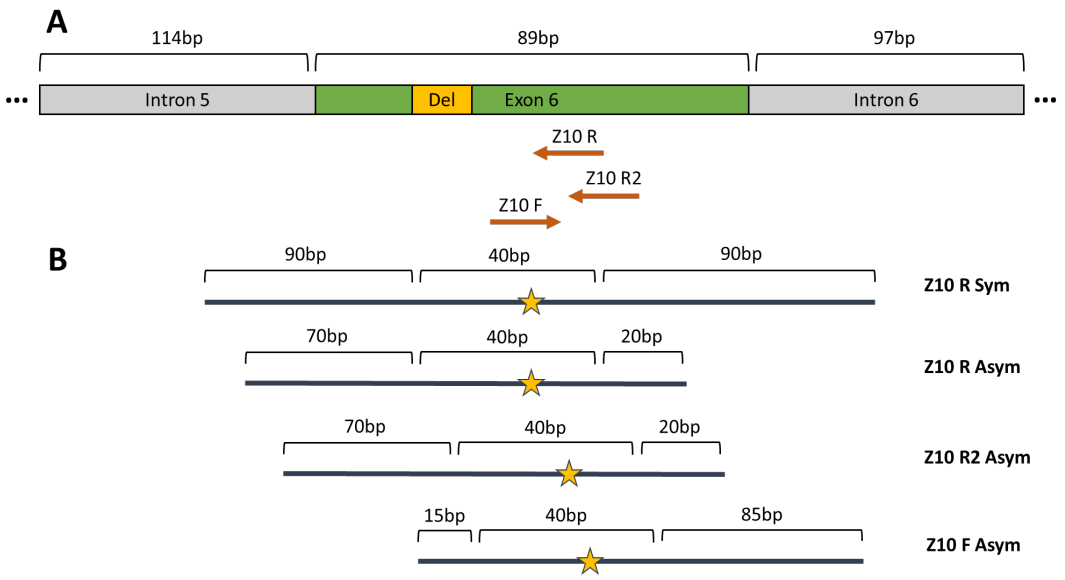


Figure 5.3: Schematic representation of the CRISPR-based repair strategy, showing guides (A) and repair templates (B) in relation to the -7bp deletion in *Zmynd10*. A) The -7bp deletion is indicated with a yellow box. Guide target sites are indicated below with orange arrows, their direction being 5' to 3' relative to the target sequence. B) ssODN repair templates specific to each guide were designed following a symmetric (Sym) or asymmetric (Asym) layout. All repair templates harbour a silent mutation which disrupts the PAM site (yellow star). Not to scale.

5.3 Generation of Mouse Embryonic Fibroblasts and Ependymal Cells from *Zmynd10* Mice

In order to test the functionality and efficiency of our various guides and repair templates, I derived two primary cell lines from the heterozygous *Zmynd10* c.695_701 line intercross; mouse embryonic fibroblasts (MEFs) from E14.5 mice (Section 2.10.1) and motile ciliated ependymal cells from E18.5 mice (Section 2.11.1). During dissection of the embryos a tissue sample was collected for genotyping. DNA was extracted from the samples (Section 2.4.1.2) and a genotyping PCR was performed using DreamTaq Green PCR Master Mix (Section 2.6) using primers 766 and 767 (Section 2.5.3). Following amplification, the PCR products were digested with MlucI (Section 2.4.4.1). As a result of the -7bp deletion, homozygous mutants lack this cut site and will show an uncleaved band at 311bp, whereas homozygous wild-type DNA will be cleaved into one 150bp and one 164bp fragment (which appears as one band on the gel). The digestion products were visualised on 2% agarose gels (Section 2.7), which indicated the presence of one homozygous mutant in the MEF E14.5 litter

(Figure 5.4A), and three homozygous mutants in the ependymal E18.5 litter (Figure 5.4B). Ependymal progenitor cells were then allowed to proliferate for 5 days prior to being used for transfections. MEFs were expanded to passage number 4 before being tested for mycoplasma and frozen down for future work (Section 2.10.1).

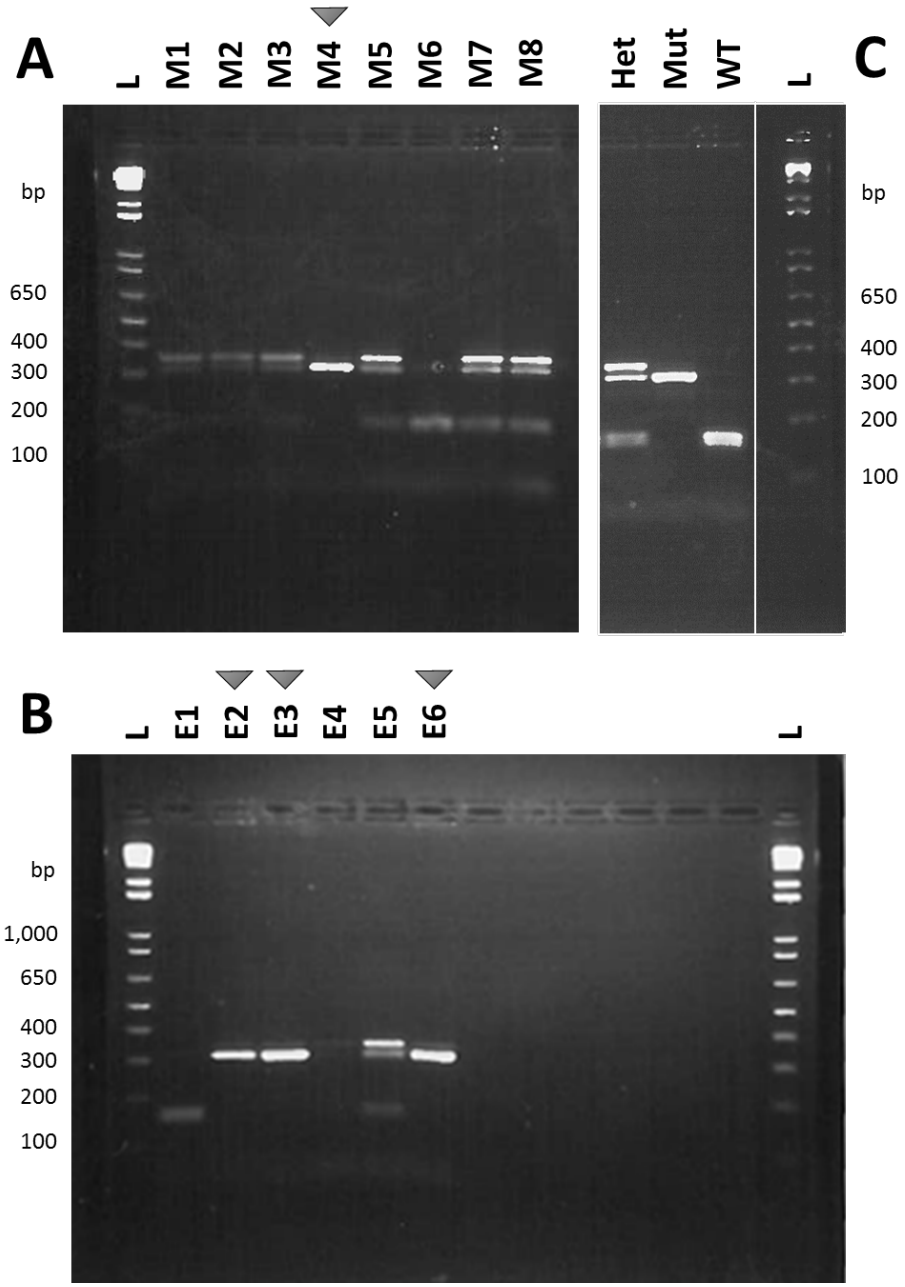


Figure 5.4: Example genotyping of MEF (A) and ependymal cell (B) progenitors, revealing the presence of homozygous -7bp deletion mutants. Genotyping PCR products were digested with MlucI and then visualised on 2% agarose gels. L = 1kb Plus DNA Ladder (Thermo Fisher). Homozygous mutants are highlighted with grey arrowheads. C) Example digestion products from a heterozygous, homozygous mutant and homozygous wild-type mouse.

5.4 Ependymal Progenitor Transfections with pX458

Following dissection, the ependymal progenitor cells were allowed to proliferate for five days to reach confluency and dissociated with TrypLE Express (Thermo Fisher). Four transfections were then performed using the Neon 10µl tips (Section 2.12.1.2):

Well	Cell Number	Genotype	Z10 R pX458 (ng)	Z10 R Sym ssODN
1	$1.5x10^5$	+/-	0	0
2	$1.5x10^5$	+/-	500	10µM, 0.75µl
3	$1.5x10^5$	-/-	500	10µM, 0.75µl
4	$1.5x10^5$	-/-	500	10µM, 1.50µl

Following transfection, the cells were returned to a 24-well plate, although all four wells appeared unexpectedly sparse. As the progenitor cells need to reach confluency to differentiate, additional non-transfected cells (of the respective genotypes) were added to wells 2 and 4. Six days post-transfection, wells 2 and 4 were observed to be confluent and had formed tight-junctions (Figure 5.5A), while wells 1 and 3 remained significantly sub-confluent (data not shown). Furthermore, GFP-positive cells could be seen in both wells (Figure 5.5B) indicating transfection had been successful. Wells 2 and 4 were therefore serum-starved to induce differentiation and the production of multi-cilia. Six days post-serum-starvation, patches of cells with motile cilia could clearly be seen in the heterozygous cells, although none overlapped with the GFP signal. GFP-positive cells, which appeared by morphology to be ependymal, were observed in the homozygous mutant well, although no cells with motile cilia were observed (Figure 5.5C).

Given these results, it was apparent that this was not the optimum system for testing guide and repair template combinations, based on the low cell numbers obtained from ependymal dissections, combined with low transfection efficiencies (Figure 5.5B, number of GFP-positive cells, not quantified) and time required for differentiation. Therefore, I decided to move onto using the *Zmynd10* mutant MEFs which are more scalable, whereupon I could assay editing at a sequence-based level (as MEFs do not produce motile cilia or machinery including ZMYND10).

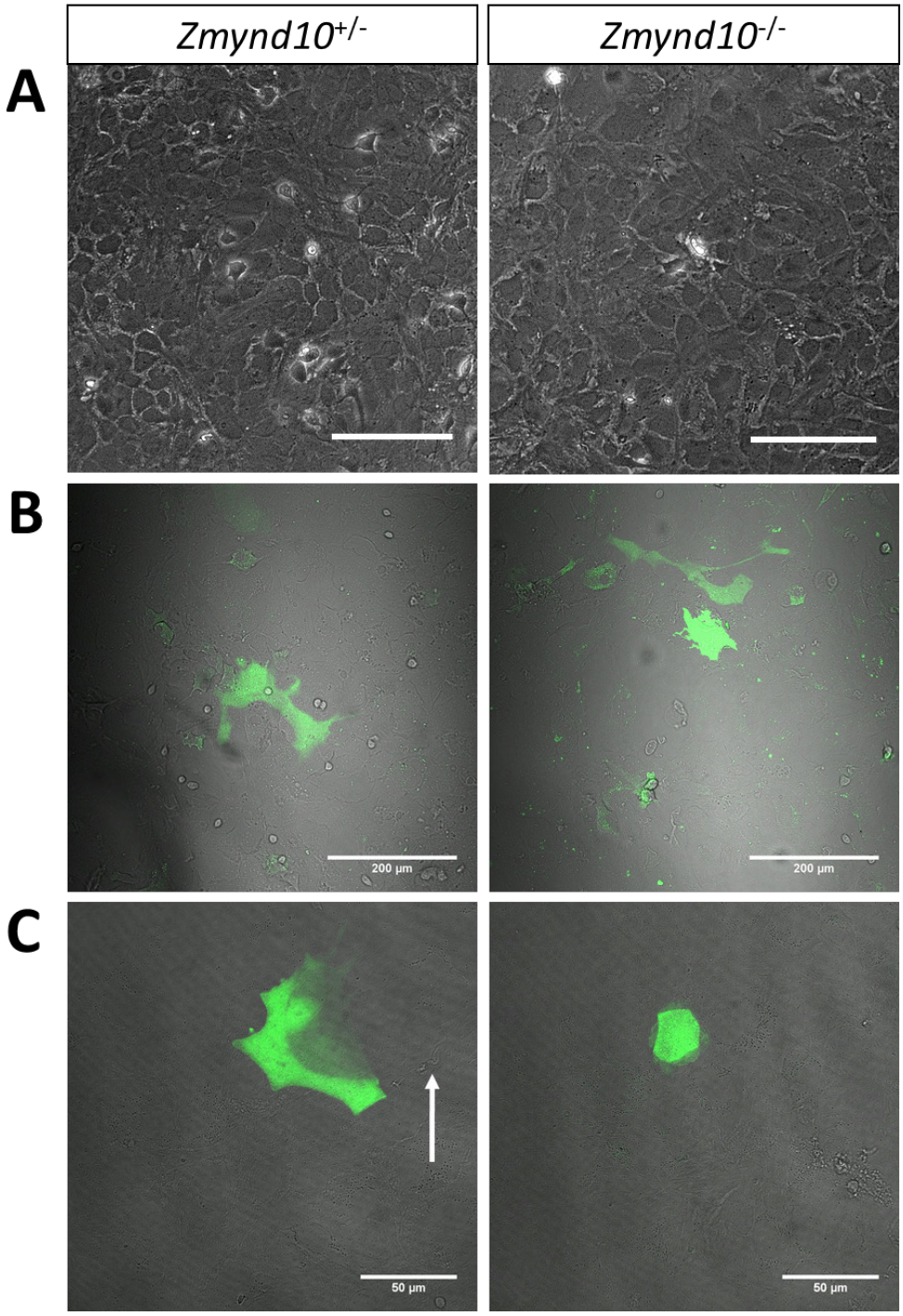


Figure 5.5: No motile cilia were observed in *Zmynd10*$^{-/-}$ ependymal cells following pX458 and ssODN repair transfection. A) Prior to serum-starvation, each well was imaged at 20x on a Nikon T-iS inverted microscope which revealed the presence of tight-junctions. Scale bar 200µm. B) Six days post-transfection, each well was imaged at 20x on a Nikon A1R confocal which revealed the presence of GFP-positive, transfected cells. C) Six days post-serum-starvation, each well was imaged at 40x on a Nikon A1R confocal which revealed the presence of motile cilia in the heterozygous population (white arrow), but not in the homozygous, transfected population. A) Brightfield, B-C) GFP + Brightfield composite.

5.5 Can Use of a Selectable GFP Transfection Marker Enrich for Editing and Repair?

As MEFs can be readily dissociated and re-plated, and can be propagated for several passages following transfection, this allows me to quantify rates of editing and repair, as opposed to transfection efficiencies, by isolating GFP-positive cells following transfection with pX458 (a similar construct to the pX330 used to generate the *Zmynd10* mice but with a GFP marker, Section 2.2.2), and then expanding and sequencing. Furthermore, as in the *mTmG* sequencing experiment, I decided to sequence the samples on the Ion Torrent Personal Genome Machine (PGM) due to its low error rate and high coverage (Quail, Smith et al. 2012), which should allow me to detect potentially rare events, such as HDR from the ssODN repair template.

5.5.1 MEF Transfections with pX458

Heterozygous and homozygous *Zmynd10* -7bp deletion MEFs were transfected using the Neon 100µl tips (Section 2.12.1.1). Two small molecules which have been reported to increase rates of HDR, BFA and L-755,507 (Yu, Liu et al. 2015), were included to test their effect in this system. Eight wells were transfected with varying conditions:

Well	Cell Number	Genotype	Plasmid	Quantity (ng)	Z10 R Sym ssODN	Small Molecule
1	$5x10^5$	+/-	N/A	0	0	N/A
2	$5x10^5$	+/-	Empty pX458	5000	0	N/A
3	$5x10^5$	+/-	Z10 R pX458	5000	0	N/A
4	$5x10^5$	+/-	Z10 R pX458	5000	10µM, 2.50µl	N/A
5	$5x10^5$	-/-	Z10 R pX458	5000	10µM, 2.50µl	N/A
6	$5x10^5$	-/-	Z10 R pX458	5000	10µM, 5.00µl	N/A
7	$5x10^5$	-/-	Z10 R pX458	5000	10µM, 2.50µl	BFA (0.1µM)
8	$5x10^5$	-/-	Z10 R pX458	5000	10µM, 2.50µl	L-755,507 (5µM)

Twenty-four hours post-transfection each well was sorted by FACS on a BD FACSJazz. No GFP-positive cells were found in well 1, as expected. Between 15,000 and 38,000 GFP-positive cells were recovered from the remaining wells, which were returned to a 24-well plate. 300,000 GFP-negative cells were also recovered from each well and returned to a 6-well plate. Seven days later, all wells were confluent and so DNA was extracted (Section 2.4.1.1) and quantified on a Nanodrop spectrophotometer

(Section 2.4.3). The DNA extractions were then named based on the well they came from, and whether they were GFP-positive/green (Gr) or GFP-negative/blank (Bl) (i.e. 5Bl represents the GFP-negative population from well 5).

5.5.2 Monitoring Genome Editing by Deep Sequencing on an Ion Torrent Platform: PCR Amplification

Sixteen barcoded Ion Torrent sequencing primers were designed to amplify the sorted populations plus a B6 (C57BL/6J) parental wild-type control (Section 2.5.4), allowing later separation of the reads by sample *in silico*. Each sample was amplified using Phusion HF Master Mix (Section 2.6) and then visualised on 1.5% agarose gels (Figure 5.6).

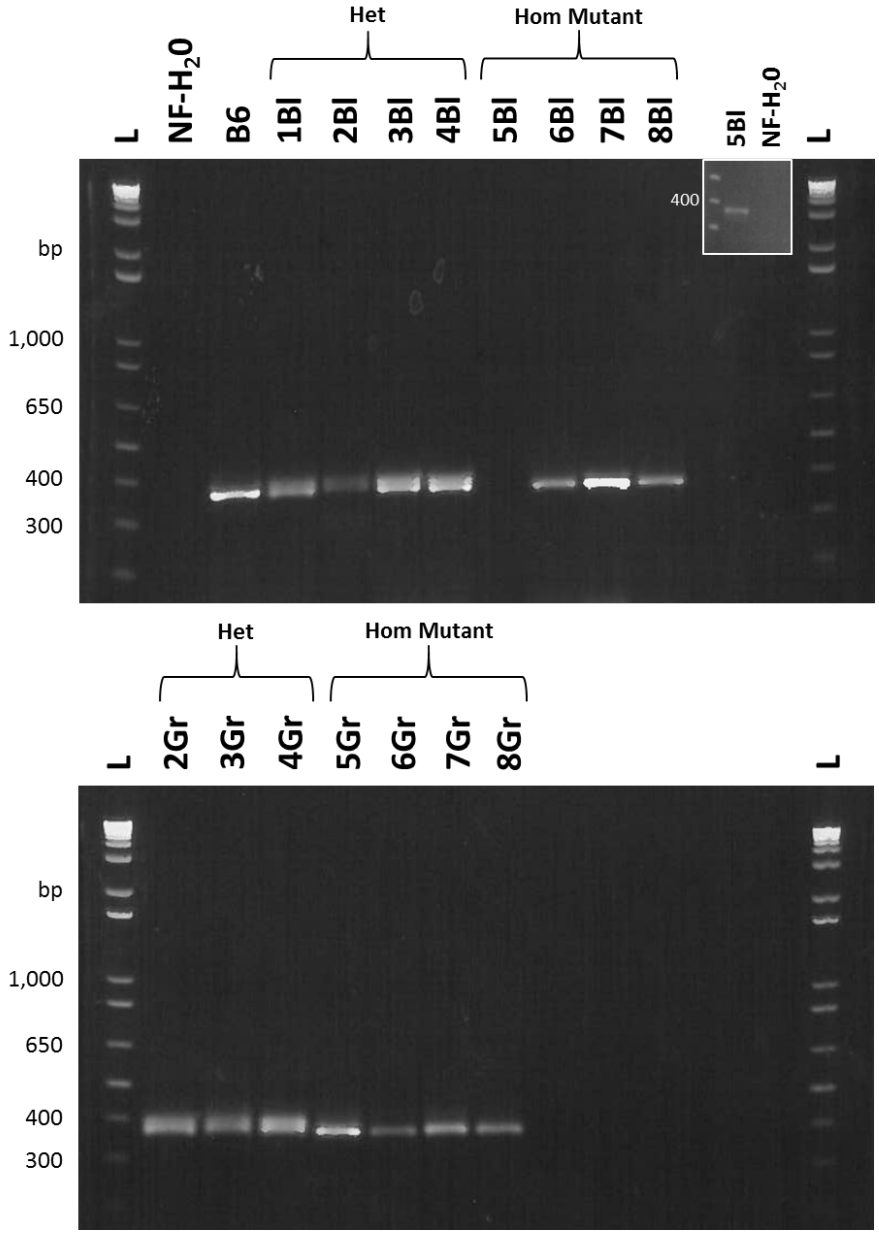


Figure 5.6: *Zmynd10* PCR amplifications for Ion Torrent sequencing. Amplifications were electrophoresed on a 1.5% agarose gel. L= 1kb Plus DNA Ladder (Thermo Fisher). Slight band blurring potentially due to issues with the buffer. Insert: amplification of 5BL initially failed due to evaporation and was repeated.

5.5.2.1 Monitoring Genome Editing by Deep Sequencing on an Ion Torrent Platform: AMPure Purification and Purity Analysis

PCR amplifications were subjected to two rounds of purification using Agencourt AMPure XP beads (Beckman Coulter) in order to remove any primer dimers present. Samples were visualised on a 1.5% agarose gel (Section 2.7) following purification which revealed that primer dimers appeared to be absent, although DNA concentration

was markedly reduced (Figure 5.7A). The purified samples were then analysed on an Agilent Bioanalyzer 2100 using a high sensitivity DNA assay which confirmed that the primer dimers had been largely removed (Figure 5.7B).

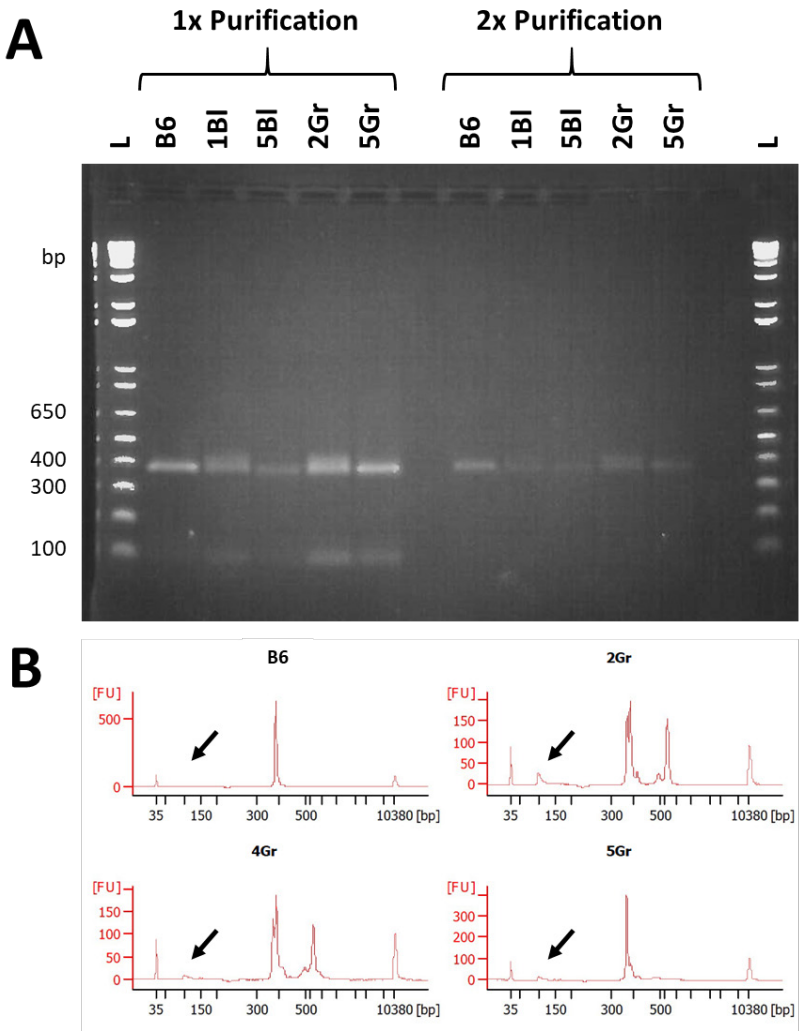
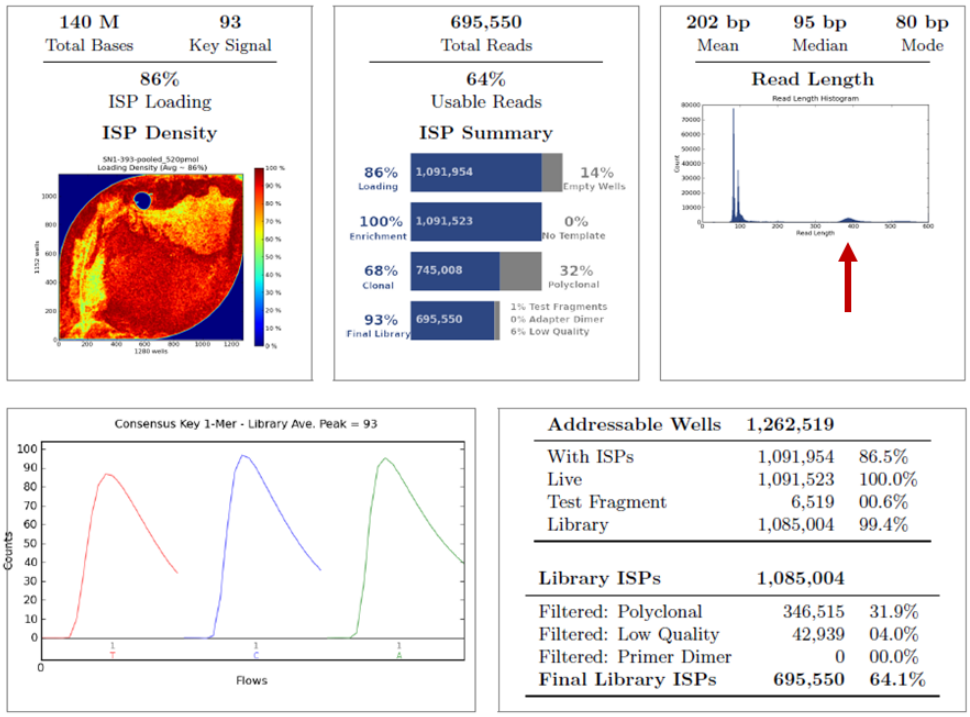


Figure 5.7: *Zmynd10* PCR products were purified with AMPure XP beads prior to Ion Torrent sequencing. AMPure XP purification leads to a substantial loss of product, although primer is also reduced. A) PCR products were electrophoresed on a 1.5% gel following one and two rounds of purification. B) Following two rounds of purification, samples were analysed on an Agilent Bioanalyzer 2100 which revealed primer dimers had been largely removed (black arrows).

5.5.3 Monitoring Genome Editing by Deep Sequencing on an Ion Torrent Platform: Ion Torrent Sequencing Results

Following purification, the samples were transferred to the IGMM technical services for sequencing on an Ion Torrent PGM, using a 314 v2 chip to obtain reads up to 400bp

in length. After sequencing was complete, a run report was generated by the PGM, which revealed that the majority of reads returned were substantially smaller than 400bp, although there was a small group of reads at this size. Unexpectedly, the run report indicated that these shorter reads were not primer dimers and did not contain any barcode sequences (Figure 5.8).



Barcode Name	Sample	Bases	≥ Q20	Reads	Mean Read Length
No barcode	none	45,329,173	14,504,979	230,333	197 bp
IonXpress_001	pooled_520pmol	4,329,642	1,628,067	20,991	206 bp
IonXpress_002	none	6,287,729	2,386,099	30,292	208 bp
IonXpress_003	none	7,566,767	2,198,390	21,022	360 bp
IonXpress_004	none	5,648,974	2,100,461	25,168	224 bp
IonXpress_005	none	6,799,906	2,669,280	35,415	192 bp

Figure 5.8: Run report generated by the Ion Torrent PGM following sequencing of MEF pX458 transfections. Not all barcodes shown. Red arrow indicates a small group of reads which are of the expected size.

5.5.3.1 The *Zmynd10* Ion Torrent Reads Have a High Error Rate which Precludes Analysis

The reads which were generated were aligned and analysed using my proprietary pipeline (Section 2.19.1). For all samples, less than 10% of the reads were found to

align to the wild-type *Zmynd10* reference sequence (ENSMUST00000010188). When viewed in the IGV browser, the reads which did align were found to have a high error rate but generally provided ~400x coverage across the -7bp deletion region. Given this error rate, however, there does not appear to be any significant enrichment in editing around the guide target site in the homozygous samples, when comparing GFP-negative (potentially non-transfected) to GFP-positive (transfected) cells. Furthermore, no reads were seen which convincingly demonstrated HDR, as all reads which lacked the deletion contained a high number of 'SNVs' (single nucleotide variants), most likely due to deteriorating quality towards the end of the reads (Figure 5.9). This high error rate appears to be due to some aspect of the in-house, IGMM Ion Torrent processing pipeline as it was not seen when *mTmG* samples were processed and sequenced by the WTCRF (Chapter 4).

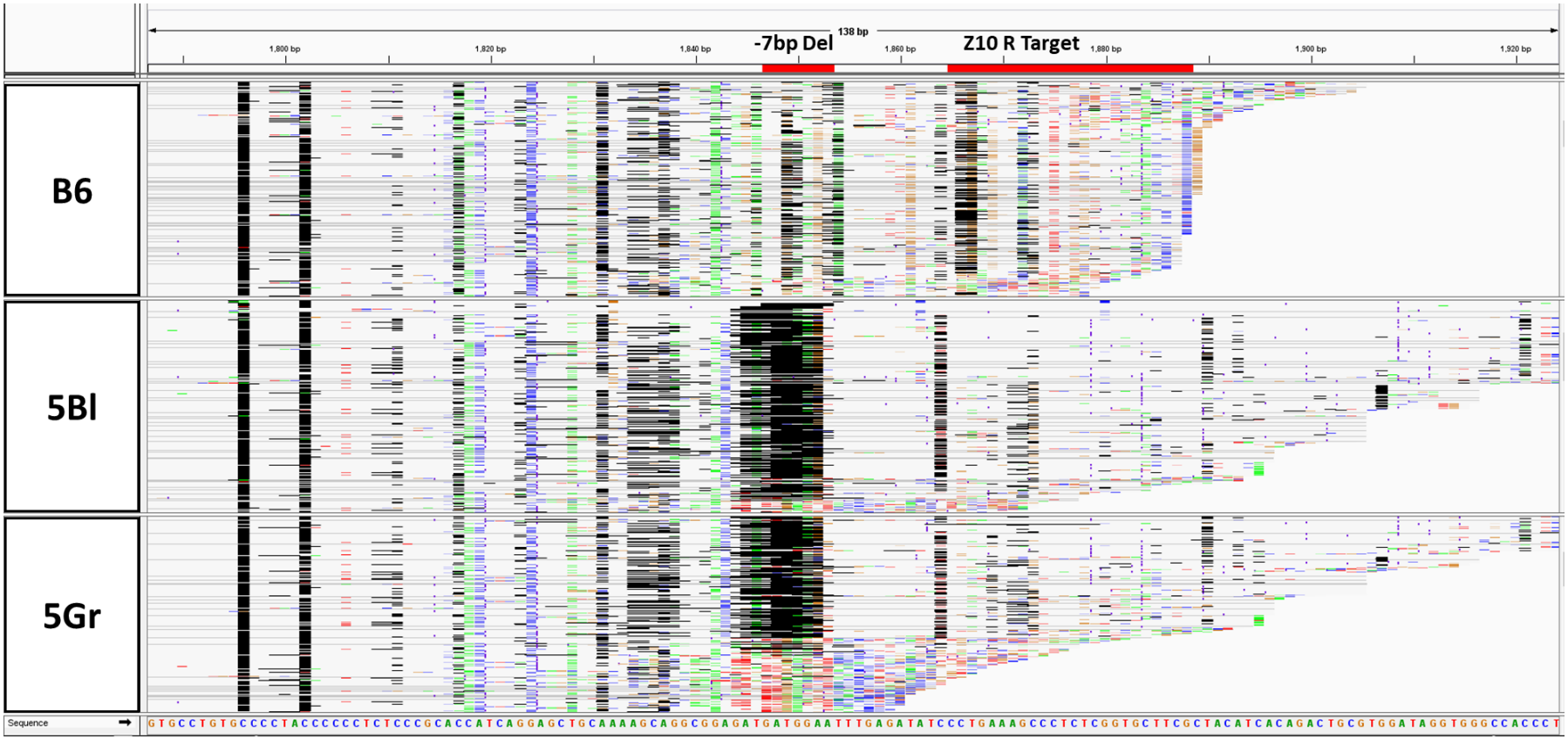


Figure 5.9: Screen capture from the IGV browser showing the alignments of Ion Torrent reads from the B6 control and a GFP-negative and GFP-positive sample against the reference, wild-type sequence for *Zmynd10*. Here, grey bars indicate uniquely aligned reads, white bars reads which aligned to multiple positions, black bars deletions and purple bars insertions. Furthermore, SNVs are indicated with coloured bars. Regions of interest have then been highlighted in red.

5.5.3.2 A Low Rate of Editing is evident in the Heterozygous, GFP-Positive Populations

Alignments of the heterozygous samples to the reference revealed several potential variants which appeared to be almost exclusively present on the wild-type, non-deletion allele, one of which (a G to C transition) may affect Z10 R guide binding (blue, Figure 5.10). Despite this, the cleaner signal in the heterozygous control sample (1Bl) allowed comparisons to be made between the editing rates in the transfected, sorted samples. This revealed that limited rates of editing above background could be detected in the GFP-positive cells, ~8%, whereas no editing could be detected in the GFP-negative cells (Figure 5.11). However, this rate of editing, post-selection, was substantially lower than expected, indicating that either the vector or the guide are highly inefficient.

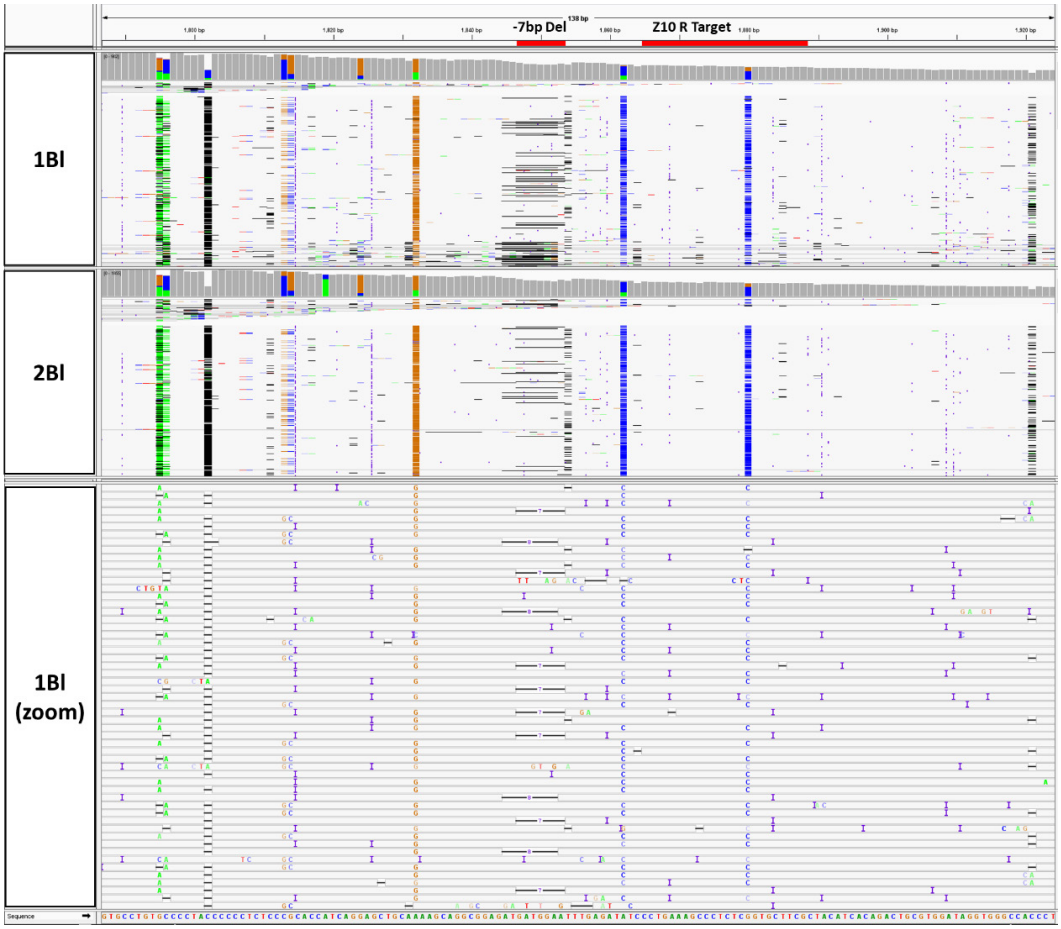


Figure 5.10: Ion Torrent alignments against the *Zmynd10* reference sequence reveal several potential variants in the heterozygous line. Screen capture from the IGV browser. Regions of interest are highlighted above in red.

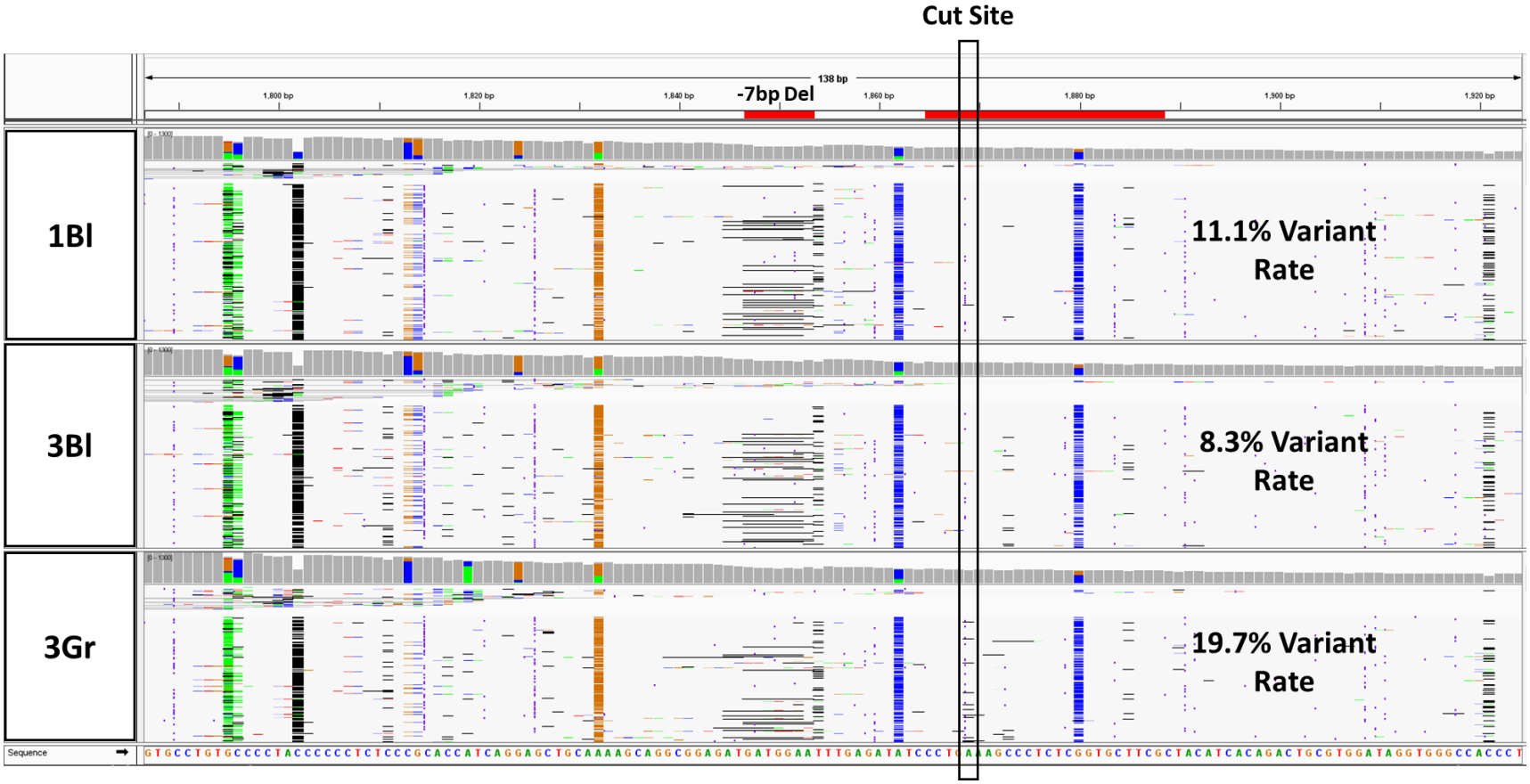


Figure 5.11: A limited rate of editing can be detected in *Zmynd10* -7bp deletion heterozygous MEFs following Z10 R pX458 transfection and sorting for GFP-positive cells. Screen capture from the IGV browser. Regions of interest are highlighted above in red. Insertions, deletions and SNVs in reads at the cut site of guide Z10 R (black box) were used to calculate an overall variant rate for each sample. The corresponding values for 4Bl and 4Gr are 8.7% and 19.1% respectively.

5.6 Sanger Sequencing of the Locus Reveals No Variants

Given the low efficiency of editing (~8%), I decided to verify that the sequence in our mouse lines matched the reference sequence (ENSMUST00000010188) which was used for guide and repair template design, as several potential variants could be seen in the Ion Torrent data. Two sets of primers were designed to sequence an ~1000bp region around the -7bp deletion (Section 2.5.5). With them, a PCR amplification was performed using Phusion HSII polymerase (Section 2.6) on three genomic DNA samples from C57BL/6J mice, BALB/cJ x C57BL/6J F1 hybrids and *Zmynd10* -7bp deletion homozygous mice. Gel electrophoresis revealed that each PCR amplification had produced a single, strong band (data not shown). The resulting PCR products were therefore submitted for Sanger sequencing by the IGMM technical services (Section 2.9). The returned sequences were aligned against the Ensembl reference sequence using the NCBI nucleotide BLAST server (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Unexpectedly, given the variability in the Ion Torrent results, Sanger sequencing revealed that all three mouse lines contain the reference sequence from Ensembl with no changes (besides the -7bp deletion), indicating our guides and repair templates should be functional.

5.7 Low Rates of Editing are Detected Post-Puromycin Selection with pX459

Given the low rates of editing seen in the Ion Torrent sequencing (~8%), along with the results of Sanger sequencing in the previous section, I concluded that the limiting factor was most likely the pX458 plasmid. Therefore, I decided to clone (Section 2.2.1) the three guides into an alternative expression vector, pX459, which is derived from the same backbone as pX330 (Section 2.2.2.3). As pX459 contains a puromycin resistance cassette, it also allows selection of successfully transfected cells, allowing us to circumvent the variable transfection efficiencies seen with pX330 (Chapter 3).

5.7.1 Puromycin Kill Curve Optimisation in MEFs

Prior to undertaking any transfections with the pX459 plasmids, I carried out a puromycin kill curve using (non-transfected) *Zmynd10* -7bp deletion homozygous and heterozygous MEFs. The results of this experiment revealed that a concentration of

6µg/ml puromycin was required to achieve 100% cell death in a non-transfected population of MEFs within 24 hours (data not shown).

5.7.2 *Zmynd10*$^{-/-}$ MEF Transfections with pX459

Six transfections were performed using the Neon 100µl tips (Section 2.12.1.1) into *Zmynd10* -7bp deletion homozygous MEFs:

Well	Cell Number	Plasmid	Quantity (ng)	ssODN	Quantity
1	$4x10^5$	Empty pX459	4000	N/A	0
2	$4x10^5$	Z10 R pX459	4000	Z10 R Sym	10µM, 4µl
3	$4x10^5$	Z10 R pX459	4000	Z10 R Asym	10µM, 4µl
4	$4x10^5$	Z10 R2 pX459	4000	Z10 R2 Asym	10µM, 4µl
5	$4x10^5$	Z10 F pX459	4000	Z10 F Asym	10µM, 4µl
6	$4x10^5$	N/A	0	N/A	0

The cells were allowed to proliferate for 24 hours post-transfection, at which point all wells had reached confluency. The media was then replaced with puromycin selection media (6µg/ml). At this time-point plasmid expression should be at its highest (DeWitt, Corn et al. 2017). After twenty-four hours incubation, the selection media was removed, at which point a small number of cells remained in the control well (well 6), although substantially less than in the other wells. All wells were then allowed to recover and proliferate for 96 hours, before DNA was extracted (Section 2.4.1.1). DNA extractions were quantified on a Nanodrop spectrophotometer (Section 2.4.3) and then diluted to ~100ng/µl. PCR amplifications were carried out (Section 2.6) using the *Zmynd10* genotyping primers (Section 2.5.3). Gel electrophoresis indicated PCR amplification had been successful, with clear bands of the expected size produced for each sample (data not shown), therefore PCR products were submitted for Sanger sequencing (Section 2.9).

5.7.3 Sanger Sequencing Reveals Negligible Rates of Editing with pX459 in MEFs

Sequence traces were inspected using the FinchTV software (Geospiza Inc.). pX459-transfected samples were compared against the no DNA, electroporation control (well 6). Sequencing revealed very limited editing with the Z10 R guide, even following puromycin selection. Z10 R2 and Z10 F appeared to elicit somewhat higher rates of

editing, although the edited sequence traces were still substantially weaker than the wild-type trace (Figure 5.12). Given the very low editing efficiencies seen, I decided not to continue with deep sequencing to attempt to identify HDR.

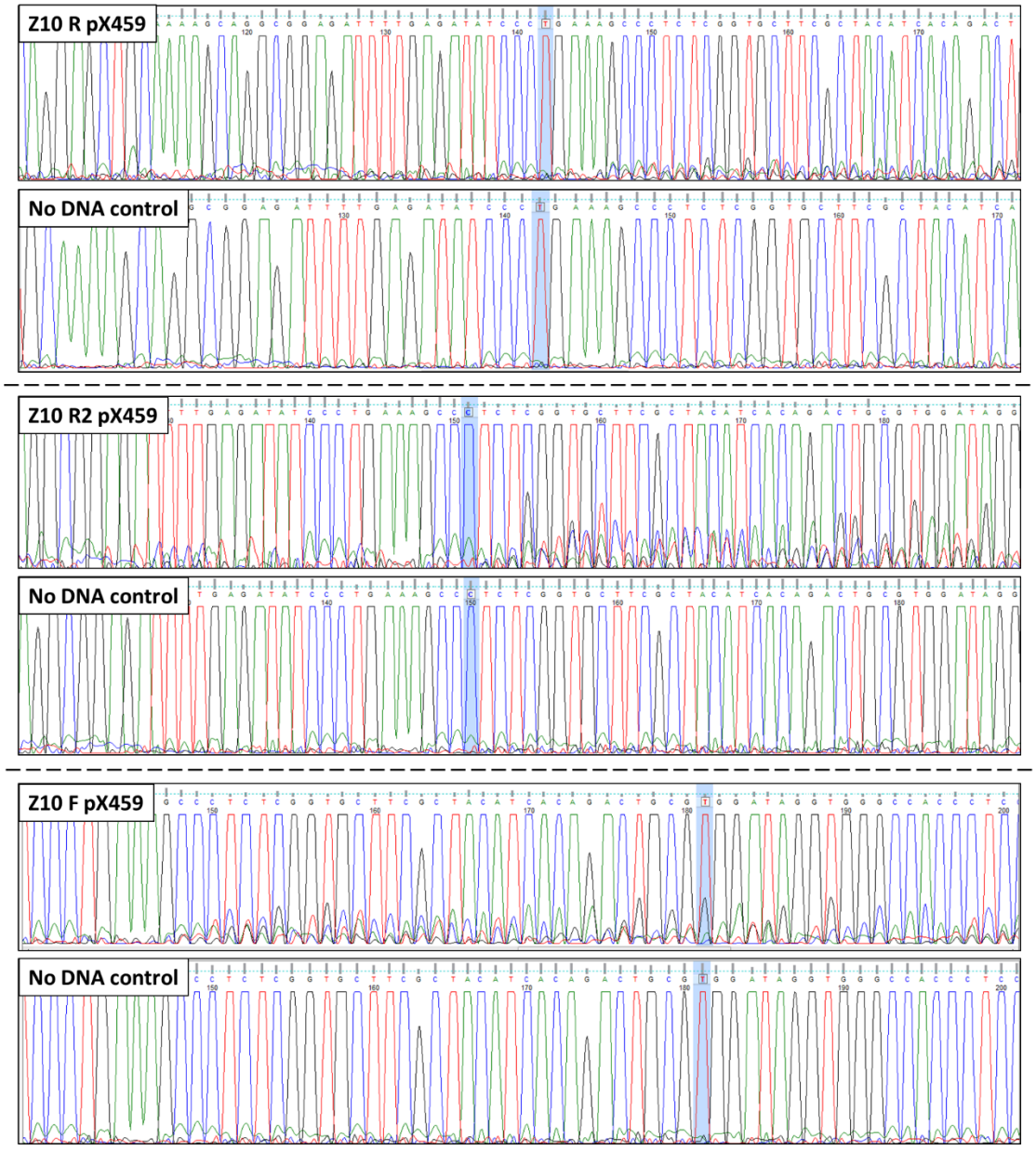


Figure 5.12: Sanger sequencing reveals very low rates of editing following pX459 transfection and puromycin selection in *Zmynd10* MEFs. Sequence traces were visualised using the Finch TV software (Geospiza Inc.). For each guide the first base of the PAM site has been highlighted in blue.

5.8 Bypassing Transcription – RNPs are Able to Drive High Rates of Editing

In this work, we have observed a significant discrepancy between the pX330 plasmids used to efficiently generate the *Zmynd10* mutant mice and the pX458/pX459 plasmids used to elicit a repair editing event, despite both containing the same editing machinery and gRNA sequences driven by the same promoter. We are unable to explain the very low efficiency of editing in positively selected cells via the pX458 or pX459 plasmids, as each clone contained sequence identical to that deposited in AddGene (www.addgene.org/). We also tested using the pX458 plasmid with different gRNAs to target another PCD candidate gene, *Wdr92*, by pronuclear injection in mice and these injections also showed no evidence of editing in any of the founders screened (data not shown). One possibility is that the P2A self-cleaving peptide between the Cas9 and selection marker leaves a small peptide fragment on Cas9 which significantly decreases its activity, particularly in mouse cells. To bypass these technical limitations, I directly delivered the same guide-RNAs together with the Cas9 nuclease protein in ribonucleotide protein complexes (RNPs). As this circumvents tissue-specific limitations on transcription and translation, it may be more efficient and comparable between cell types. Furthermore, I had previously seen high rates of editing (>30%) in my own hands using RNP complexes in the *mTmG* MEFs (Figures 3.36, 3.37).

5.8.1 *Zmynd10*$^{-/-}$ MEF RNP and ssODN Repair Transfection

Guide-RNAs were synthesised for each target sequence (Section 2.2.3), complexed with GeneArt Platinum Cas9 Nuclease (Thermo Fisher), and electroporated into *Zmynd10* -7bp homozygous MEFs along with the ssODN repair templates (Section 2.12.1.3):

Well	Cell Number	Z10 R2 RNP (ng)	Z10 F RNP (ng)	ssODN	Quantity	Cas9 (µg)
1	$3x10^5$	0	0	N/A	0	6
2	$3x10^5$	1,440	0	Z10 R2 Asym	10µM, 3µl	6
3	$3x10^5$	0	1,440	Z10 F Asym	10µM, 3µl	6

I decided to focus on guides Z10 R2 and Z10 F as they had shown clearer evidence of editing with the pX459 plasmid (Figure 5.12). Forty-eight hours post-transfection DNA was extracted from all wells (Section 2.4.1.1) and PCR amplifications were performed (Section 2.6) using the *Zmynd10* genotyping primers (Section 2.5.3). PCR amplifications were then purified (Section 2.4.2).

5.8.2 Surveyor and Restriction Enzyme Digest Results Are Inconclusive

The purified PCR products were digested with MlucI (Section 2.4.4.1) to attempt to detect repair. The -7bp deletion removes this cut site, which is exactly mid-way through the PCR product, and hence homozygous mutants will show a single, 314bp band. Repaired alleles will then be detected by their smaller size post-digestion, at 150bp and 164bp. Furthermore, a Surveyor assay was carried out (Section 2.4.4.2) to estimate overall rates of mutagenesis. Purified PCR products, MlucI digests and Surveyor products were visualised on a 2% agarose gel (Figure 5.13). The gel revealed that the PCR and purification were successful, although no bands indicative of HDR could be seen. The Surveyor assay indicated that mutagenesis may have occurred via the reduced intensity of the main band at ~310bp in wells 2 and 3, although any lower bands were very faint, suggesting a large range of products may have been produced. In order to gain a better understanding of the rates of mutagenesis, all three PCR products were submitted for Sanger sequencing (Section 2.9).

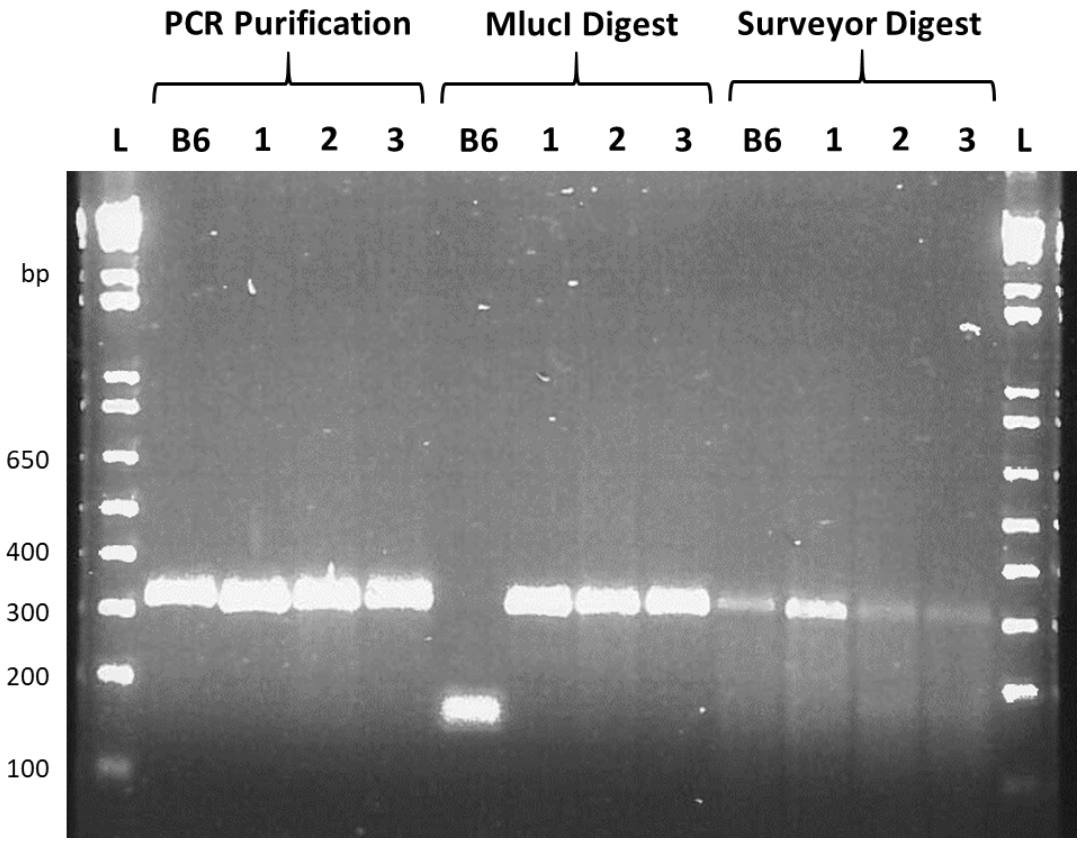


Figure 5.13: Surveyor and restriction enzyme digest results are inconclusive following RNP transfection into *Zmynd10*$^{-/-}$ MEFs. Products were electrophoresed on a 2% agarose gel. L= 1kb Plus DNA Ladder (Thermo Fisher).

5.8.3 Sanger Sequencing Reveals a High Frequency of Editing

Sequence traces were visually inspected using FinchTV (Geospiza Inc.), revealing a substantial decrease in wild-type signal 3' of the cut sites for both guides, suggesting >50% rate of mutagenesis (Figure 5.14). Based on this, I decided to TOPO-clone both PCR products to attempt to identify alleles which had undergone HDR.

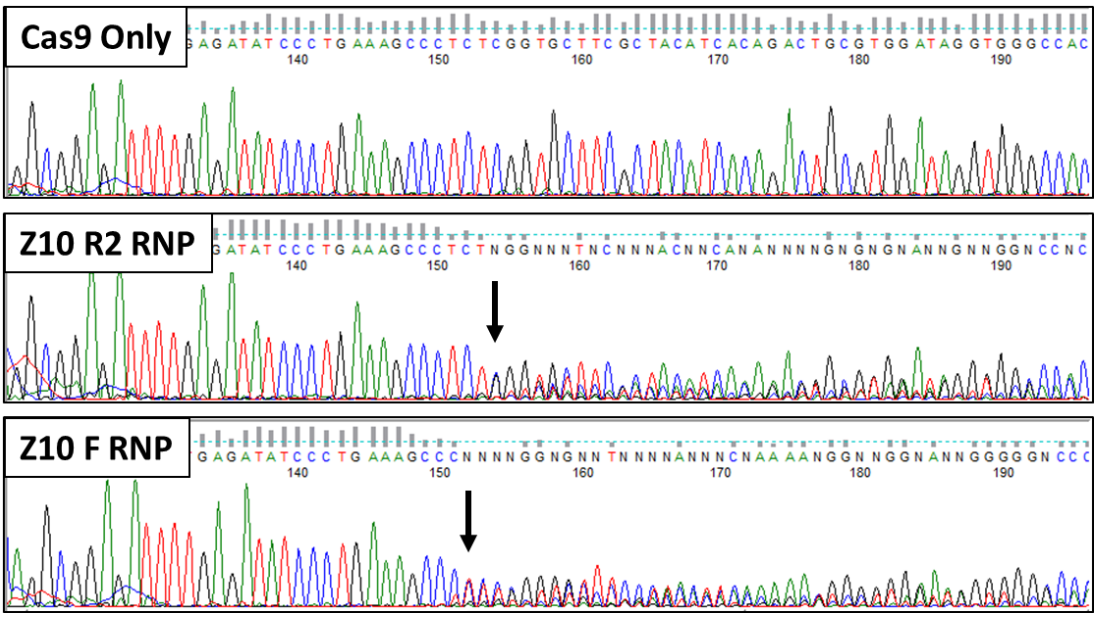


Figure 5.14: Sanger sequencing reveals a high rate of editing following RNP transfection into *Zmynd10*$^{-/-}$ MEFs. Sequence traces were visualised using the FinchTV software (Geospiza Inc.). Cut sites for each guide have been indicated (black arrow).

5.8.4 TOPO-Cloning Reveals a Very High Rate of NHEJ Following RNP Transfection with Evidence of Partial Repair

The purified PCR products for Z10 R2 and Z10 F were TOPO-cloned (Section 2.8), and 96 clones from each were submitted for Sanger sequencing (Section 2.9). Sequences were then aligned against the *Zmynd10* reference sequence using SnapGene (GSL Biotech LLC) and editing outcomes were manually quantified for each guide (Figure 5.15A). This analysis revealed very high rates of editing following RNP transfection with both guides, at ~80%. Furthermore, the alignments revealed one clone (transfected with Z10 R2 RNP and Z10 R2 Asym ssODN) which had undergone HDR, repairing the -7bp deletion and mutating the PAM site (Figure 5.15B, clone 31). Unexpectedly, however, this clone contained a novel deletion 3' of the -7bp deletion site, despite the fact that the mutated PAM site should have abolished cleavage. We hypothesised that this may be due to microhomology at the locus or in the repair template, however microhomology prediction (Bae, Kweon et al. 2014) failed to identify the residues surrounding the novel deletion. The existence of errors following repair, even with PAM-blocking mutations, has been documented, although it is relatively infrequent (two- to ten-fold less than 'scarless' repair). It is hypothesised it

could be due to prior or concomitant use of the NHEJ pathway (Paquet, Kwart et al. 2016).

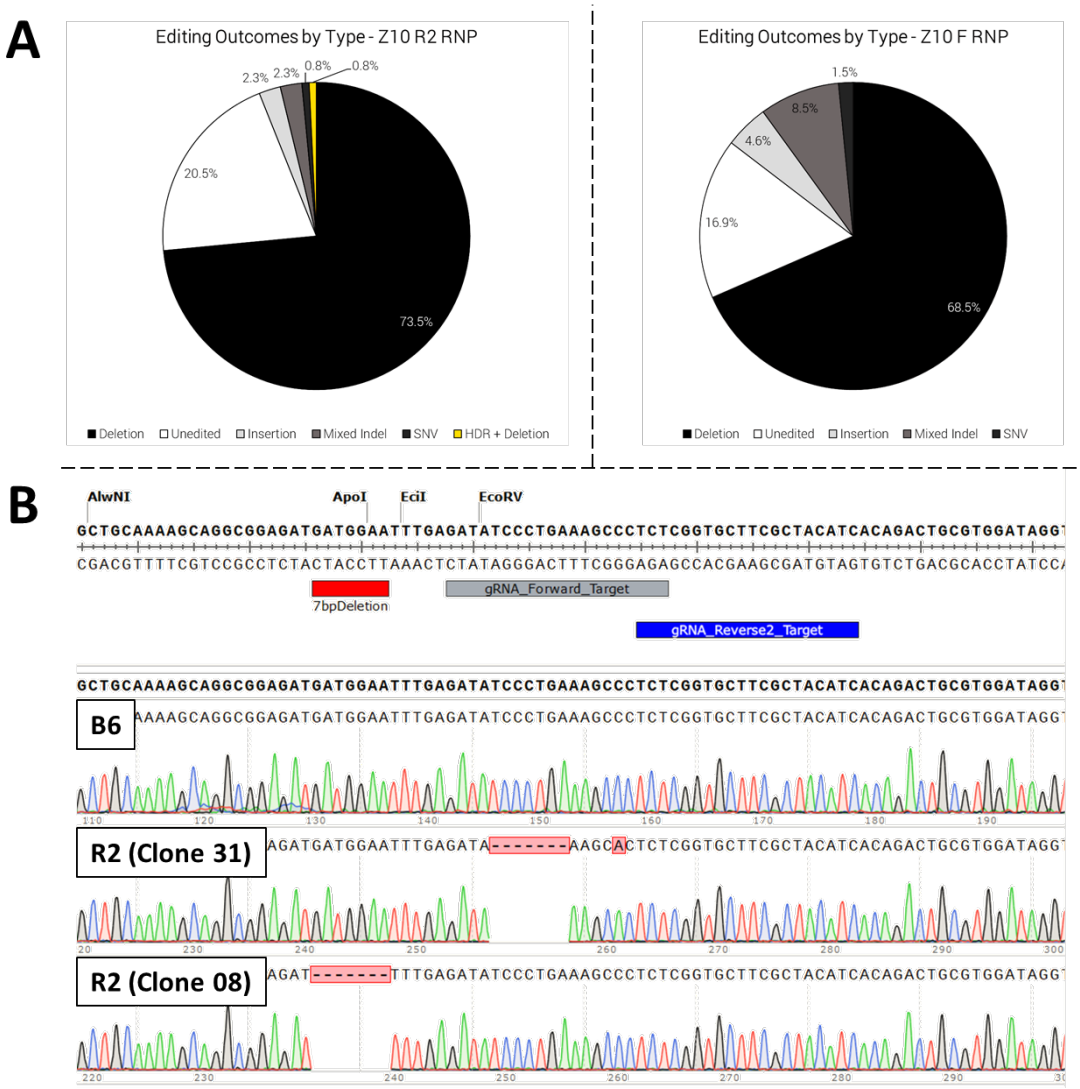


Figure 5.15: TOPO-cloning reveals an editing rate of ~80% following RNP transfection in *Zmynd10*$^{-/-}$ MEFs. A) Sequences from TOPO-cloning were aligned against the wild-type reference for *Zmynd10* using SnapGene and alignments were manually quantified for editing. B) SnapGene alignments revealed one clone transfected with Z10 R2 (Reverse2) and ssODN repair (clone 31) which had undergone HDR yet carried a novel deletion between the repair site and the mutated PAM.

5.9 Is the *mTmG* System a Good Predictor of Editing at Other Loci? - Dual Transfections into *mTmG* MEFs

As we had found a method, RNP transfection, which can generate high rates of editing at both the *Zmynd10* -7bp deletion locus and visually in the *mTmG* system, we wanted to investigate whether there is a correlation between editing at the two loci. If so, this would provide preliminary evidence that an individual cell's permissiveness for editing by CRISPR is similar across the genome, and thus that we can extrapolate results gained with the *mTmG* system to alternative, disease-causing loci. In order to test this, immortalised heterozygous *mTmG* MEFs were transfected with both LoxP#1 RNP (Section 3.16.2) and Z10 R2 RNP complexes simultaneously using the Neon 10µl tips (Section 2.12.1.3):

Well	Cell Number	LoxP#1 RNP (ng)	Z10 R2 RNP (ng)	Cas9 (µg)
1	$0.5x10^5$	0	0	1
2	$0.5x10^5$	240	0	1
3	$0.5x10^5$	120	120	1

Seven days post-transfection each well was imaged at 20x on a Nikon A1R confocal microscope, which revealed a high number of GFP-positive cells in both well 2 and 3 (Figure 5.16A). Therefore, all three wells were sorted on a BD FACSJazz, separating the visually edited cells (Q2-Q4) from the Tomato-only population (Q1) (Figure 5.16B). Furthermore, an unsorted population was kept for comparison. DNA extractions were then taken (Section 2.4.1.1) and PCR amplifications were performed (Section 2.6) using the *Zmynd10* genotyping primers (Section 2.5.3) which produced a single band for each sample (Figure 5.16C).

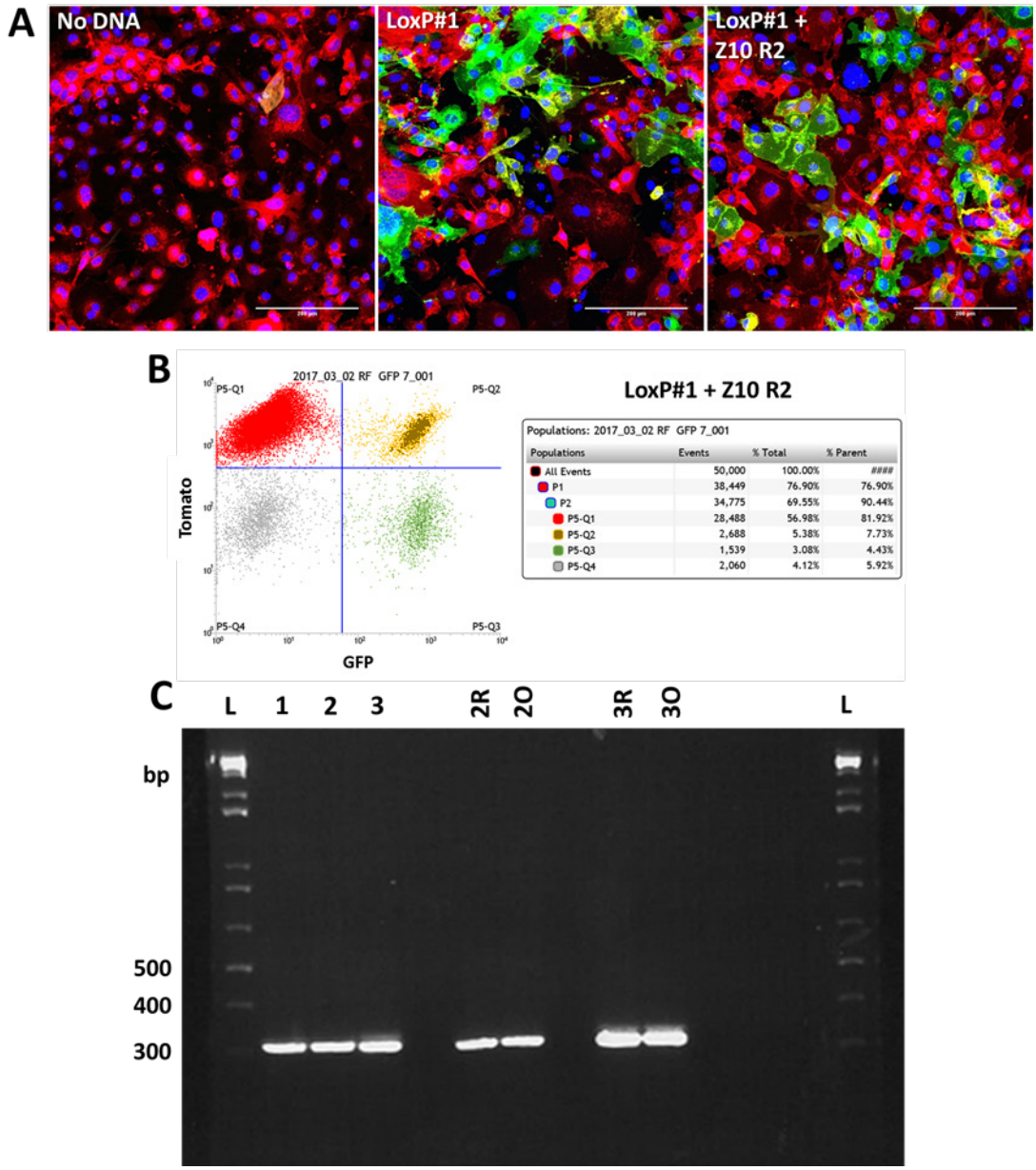


Figure 5.16: Dual transfections with *Zmynd10* and *mTmG* RNP complexes in heterozygous *mTmG* MEFs. A) Seven days post-transfection each well was imaged at 20x. Scale bar 200µm. B) Eight days post-transfection each well was sorted, separating Tomato-only cells in Q1 (top-left quadrant) from visually edited cells in Q2-4. C) PCR amplifications were performed across the *Zmynd10* -7bp deletion locus in sorted and unsorted populations and products were visualised by gel electrophoresis on a 2% agarose gel. Here R = Q1 and O = Q2-4 on the FACS plot. L = 1kb Plus DNA Ladder (Thermo Fisher).

PCR amplifications were submitted for Sanger sequencing (Section 2.9), and sequence traces were then visually inspected using FinchTV. Unexpectedly, no editing could be detected at the *Zmynd10* locus in the visually edited *mTmG* MEFs (Figure 5.17A), however there appeared to be a SNV present which may affect guide binding. Therefore, the sequence from the non-transfected (heterozygous *mTmG*) control was aligned to the *Zmynd10* reference sequence using nucleotide BLAST (NCBI). This revealed the presence of two variants in our *mTmG* line (Figure 5.17B), which is on an out-bred rather than B6 background. One of these variants is located proximal to the PAM site for Z10 R2 (Figure 5.17A, arrow) and also disrupts the PAM site for Z10 F, therefore, cleavage rates would be expected to be substantially reduced with both guides. If cleavage did occur, the locus may be repaired based on the sequence of the alternative allele, complicating any analysis.

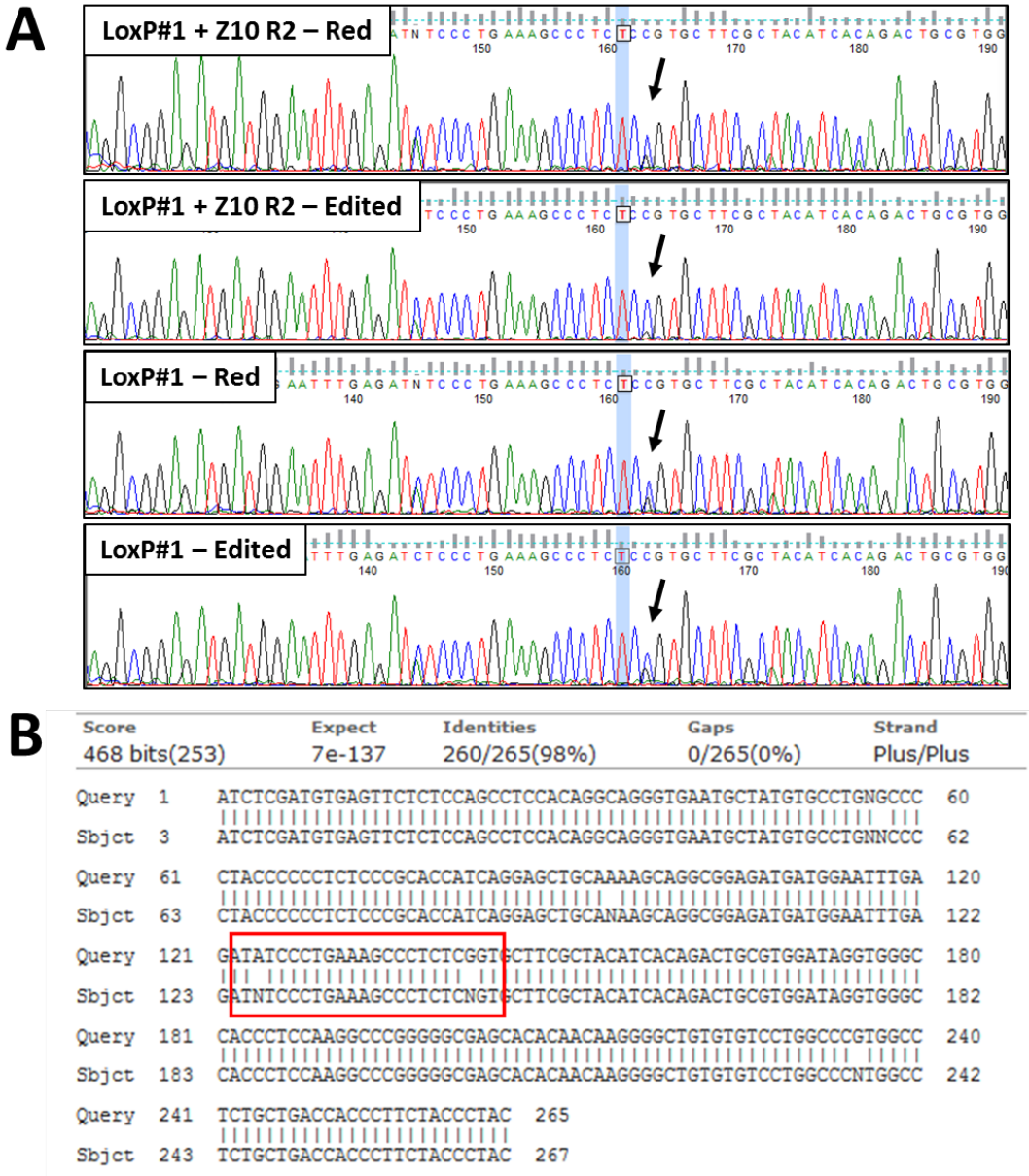


Figure 5.17: No editing was detected at the *Zmynd10* -7bp deletion locus in the heterozygous *mTmG* MEFs due to the presence of SNVs. A) Sequence traces were visualised using FinchTV (Geospiza Inc.). No editing could be seen at the *Zmynd10* locus in the samples transfected with Z10 R2, however a SNV (black arrow) may have affected binding. The predicted cut site for Z10 R2 has been highlighted in blue. B) A nucleotide BLAST (NCBI) alignment of the heterozygous *mTmG* sequence (Sbjct) against the reference sequence (Query) reveals two SNVs (red box) which will affect binding of both *Zmynd10* guides. These two SNVs were also present and variable in homozygous *mTmG* MEFs (data not shown).

5.10 Conclusions and Future Work

The results of this chapter highlight the inherent variability in the genome, which can be substantial even within species. Therefore, it is prudent to sequence across the locus one wishes to target in one's strain of interest *prior* to designing any reagents for genome editing. The implications of this also carry through to a therapeutic context, wherein it may be necessary to sequence each patient before designing a bespoke intervention. Furthermore, Cas9 expression vectors may have markedly different efficiencies in different cell types, potentially due to the promoter employed. Alternatively, the direct delivery of Cas9 in RNP complexes may offer high rates of editing in many cell types without need for such considerations. Moreover, the short ~24 hour residency time of RNPs may reduce the frequency of off-target editing (DeWitt, Corn et al. 2017), which is critical in a therapeutic context.

In the future, it will be necessary to repeat the *Zmynd10* RNP and repair transfections in MEFs, potentially employing a plasmid repair template whose longer regions of homology may increase HDR (Hendel, Kildebeck et al. 2014, Li, Wang et al. 2014). If repair were detected, it would then be apt to employ the same conditions in a motile ciliated cell type such as ependymal cells, and assay whether a) the sequence can be corrected and b) whether repair at the *Zmynd10* locus leads to functional correction in terms of restoring ciliary motility. If this is the case, it would be exciting to attempt to employ these reagents *in vivo* in the *Zmynd10* mouse model at several different stages (e.g. blastocyst or post-natal) and determine whether the PCD phenotype can similarly be rectified. Furthermore, it will be necessary to repeat the experiment in the *mTmG* MEFs using novel guides which target regions unaffected by SNVs. If there is a correlation between editing at the *mTmG* locus and the *Zmynd10* locus, this will have far-reaching implications about editing efficiencies between ubiquitously expressed loci, such as *Rosa26*, and disease loci, such as *Zmynd10*, with highly tissue-specific expression profiles.

Chapter 6 Concluding Remarks

The allure of gene therapy lies in its potential, theoretically, to correct the underlying causes of all genetic diseases, whether inherited or acquired. However, despite being initially postulated as early as 1970, we still face many of the issues discussed in the first comprehensive review by Friedmann and Roblin (1972); such as efficient delivery of the DNA, targeting of the correct cell types, preventing an immune response, and ensuring correct regulation. Recent work with delivery vectors, in particular AAV (Baruteau, Waddington et al. 2017, Gray-Edwards, Randle et al. 2017, Guggino and Cebotaru 2017, Wells 2017, Yu, Mookherjee et al. 2017), is addressing the first three concerns, while SSN-mediated genome editing provides us with a novel, potentially safer, modality which, via HDR, should assist with ensuring correct regulation. To date, however, there remain no good models to test and optimise gene therapy by gene correction, in particular with the revolutionary CRISPR/Cas9 system (Stella and Montoya 2016).

Given this, the *mTmG* reporter system we have developed should provide invaluable preclinical results regarding *in vivo* editing and repair driven by CRISPR/Cas9 and allow optimal reagents and protocols to be swiftly developed for the targeting of any tissue. To this end, I foresee that this system could also be repurposed, with appropriate imaging or FC/FACS technology, to undertake small molecule screens to identify novel enhancers of HDR, in various tissue-specific cell lines. However, further work needs to be done in comparing the editing seen at the *mTmG* locus with other loci, to determine how broadly applicable the results are. The advent of single cell sequencing (Huang, Ma et al. 2015) will allows us to comprehensively demonstrate simultaneous editing at the *mTmG* locus and disease-causing loci, while single cell RNA sequencing (Kolodziejczyk, Kim et al. 2015) will allow us to probe the effects of any editing achieved on a cell-by-cell basis, including, for example, the induction of DNA damage or repair proteins; providing important mechanistic insights.

Before CRISPR/Cas9 can be applied therapeutically, however, certain concerns must be addressed. Chief of these is the prospect of off-target editing leading to adverse outcomes, in particular oncogenesis. Whilst it appears as if off-target mutagenesis can generally be predicted and occurs at low levels in cell lines (Section 1.5.1), due to

inherent genomic variability in a natural population it will be harder to predict the on- and off-target efficiencies of a certain guide sequence with accuracy. Indeed, each individual in the 1,000 genomes project contains 4 to 5 million variants relative to the reference (Lessard, Francioli et al. 2017). Prior to undertaking any work, it is therefore wise to carefully profile the off-target sites of chosen guides and discard any which have <5 mismatches in and around known oncogenes. Furthermore, with the reducing cost of NGS (Weymann, Laskin et al. 2017), it may become feasible to undertake WGS in each patient prior to designing a targeting strategy. Alternatively, novel methods of detecting and sequencing CRISPR/Cas9-mediated DSBs (Tsai, Zheng et al. 2014, Tsai, Nguyen et al. 2017) may allow a preclinical evaluation of each guide on a patient-by-patient basis. However, given the incredibly high, and increasing, specificity of novel, engineered Cas9 variants (Chen, Dagdas et al. 2017, Casini, Olivieri et al. 2018), worries of off-target mutagenesis may soon be laid to rest.

Focussing specifically on gene therapy for PCD, several preliminary investigations have been made into both gene replacement approaches and gene editing/correction methodologies. In terms of conventional gene replacement, three studies have been carried out. Two of these studies were able to demonstrate at least partial restoration of ciliary function in cell culture assays; one used an SIV vector pseudotyped with VSV-G (vesicular stomatitis virus glycoprotein G) to transduce DNAI1 cDNA into human airway epithelial cells (Chhin, Negre et al. 2009) and the other used a novel lentiviral vector pseudotyped with avian influenza hemagglutinin to apically transduce mouse tracheal epithelial cells with Dnaic1 cDNA and achieving ~10% restoration of ciliary function (Ostrowski, Yin et al. 2014). Interestingly, this second study also used the same vector *in vivo* to transduce control and PCD mice by nasal inhalation and found a high level of gene transfer to control animals but little/no transfer to PCD mice; likely due to the inhibitory effect of mucus (Ostrowski, Yin et al. 2014). Conversely, a separate study has shown a significant improvement in the ciliopathy phenotype of ORPK (Oak Ridge Polycystic Kidney) mice, who display anosmia and a failure to thrive, following intranasal injection with IFT88–GFP adenovirus. Transduced ORPK mice displayed a restoration of ciliary structure on the olfactory neurons and an ~60% increase in weight at 3 weeks of age (McIntyre, Davis et al. 2012); although in this model mucus plugs are likely not an issue.

Many of the genes which are currently known to cause PCD, however, are much too large to fit into commonly used gene replacement vectors (e.g. DNAH11 is ~14kb (Lai, Pifferi et al. 2016) versus a maximum packaging capacity of ~8kb for lentivirus and adenovirus (Boink and Robinson 2014)). Gene editing approaches therefore allow the field of PCD gene therapy to sidestep these packaging constraints by specifically correcting the mutated base(s) in the patient's genome. One study has investigated this approach so far. In this study airway epithelial cells were collected from two patients with nonsense mutations in DNAH11. These cells form spheroids which do not rotate in culture. However, following transduction with vectors encoding left and right TALENs and a wild-type repair template, ~30% normalisation of ciliary activity was seen in spheroids from both patients (Lai, Pifferi et al. 2016); a potentially therapeutically relevant proportion of cells. To investigate the level of correction needed, a separate study, which utilised a floxed Dnaic1 gene to mosaically delete Dnaic1 in the airways of mice, demonstrated that mucociliary clearance began to be detected at low levels when Dnaic1 protein levels rose above 15% in some animals (Ostrowski, Yin et al. 2014); a seemingly achievable target.

However, for gene therapy to progress to the clinic, several concerns must be addressed which are pertinent to both gene replacement and gene editing strategies. These concerns chiefly regard immunogenicity; of the delivery vector, of the nuclease employed (if any), and of the protein of interest which has been reintroduced. Much work has already been done on engineering delivery vectors to increase their specificity to their target cells and reduce any potential immune response (e.g. Martino, Basner-Tschakarjan et al. (2013)**,** Lisowski, Dane et al. (2014)), while any immune response to the nuclease could be reduced by employing transient expression or inducible systems (Cao, Wu et al. (2016), Liu, Ramli et al. (2016), Lu, Zhao et al. (2018), Senturk, Shirole et al. (2017)) and/or tagging the nuclease for degradation (Tu, Yang et al. 2017). Regarding the potential for an immune response to the reintroduced protein (which may never have previously been present in its entire, correctly folded form), a recent study demonstrated expansion of corrected hepatocytes containing wild-type Fah protein *in vivo* following CRISPR-Cas9-mediated repair; indicating no substantial immune response to the protein which was only previously present in a severely truncated, and then degraded form (Yin, Xue et al. 2014). However, this may

need to be investigated on a gene-by-gene, and by tissue basis. Furthermore, oncogenesis is always a concern if an integrating vector (such as lentivirus) is employed, however, it is possible to profile via sequencing the integration sites of the vector thoroughly prior to moving into the clinic (Afzal, Wilkening et al. 2017), and it may also be possible to direct vectors specifically to safe-harbour sites (Rio, Banos et al. 2014).

Finally, in order for any treatment to be of significant benefit, it has to be accessible to patients. The first commercialised gene therapy treatment, Glybera (alipogene tiparvovec), an AAV-mediated approach for the ultra-rare disorder lipoprotein lipase deficiency (European Biotechnology 2017), appears to have failed at the first hurdle. With its incredibly high cost, $1,000,000 per treatment, and limited efficacy, it was withdrawn from the market in the US after being given to only one patient, despite its safety profile. Furthermore, Glybera was not recommended for national reimbursement in any European country (Senior 2017). The second gene therapy treatment to be approved, Imlygic (talimogene laherparepvec), for melanoma, has currently fared little better. In the UK, the only country where it has currently been assessed, it was deemed not cost-effective and in need of greater supporting evidence (Touchot and Flume 2017).

These initial forays into commercialising gene therapy highlight the need for a product which delivers substantial improvements in terms of its clinical end-points and a suitable pricing model. There is, however, a recent success story in Strimvelis (GlaxoSmithKline), an *ex vivo, gammaretrovirus*-based stem cell gene therapy for ADA-SCID (severe combined immunodeficiency due to adenosine deaminase deficiency). Since 2000, over 40 patients have been treated with a 100% survival rate, 75% disease-free (Booth, Gaspar et al. 2016). It was approved in Europe in 2016, and the Italian Medicines Agency, AIFA, has agreed to reimburse the €594,000 cost. Critically, the cost of Strimvelis treatment is less than lifelong enzyme replacement therapy for ADA-SCID patients (Touchot and Flume 2017). Therefore, Strimvelis should provide a guiding light for the many gene therapies currently in clinical trials. Furthermore, the uptake of highly modular, ‘personalised’ gene correction strategies using CRISPR/Cas9 should only reduce costs in safety testing and increase

efficiencies, potentially bringing about the revolution of gene therapy first discussed almost 50 years ago.

List of References

Afzal, S., S. Wilkening, *et al.* (2017). "GENE-IS: Time-Efficient and Accurate Analysis of Viral Integration Events in Large-Scale Gene Therapy Data." Molecular Therapy. Nucleic Acids **6**: 133-139.

Ahnesorg, P., P. Smith, *et al.* (2006). "XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining." Cell **124**(2): 301-313.

Alton, E. W., J. M. Beekman, *et al.* (2017). "Preparation for a first-in-man lentivirus trial in patients with cystic fibrosis." Thorax **72**(2): 137-147.

Alton, E. W. F. W., D. K. Armstrong, *et al.* (2015). "Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial." The Lancet Respiratory Medicine **3**(9): 684-691.

Ambardar, S., R. Gupta, *et al.* (2016). "High Throughput Sequencing: An Overview of Sequencing Chemistry." Indian J Microbiol **56**(4): 394-404.

Amoasii, L., C. Long, *et al.* (2017). "Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy." Sci Transl Med **9**(418).

ATUM. (2018). "CRISPR gRNA Design Tool." Retrieved 18th February, 2015, from https://www.atum.bio/eCommerce/cas9/input.

Bae, S., J. Kweon, *et al.* (2014). "Microhomology-based choice of Cas9 nuclease target sites." Nat Methods **11**(7): 705-706.

Bae, S., J. Park, *et al.* (2014). "Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases." Bioinformatics **30**(10): 1473-1475.

Banasavadi-Siddegowda, Y. K., J. Mai, *et al.* (2011). "FKBP38 Peptidylprolyl Isomerase Promotes the Folding of Cystic Fibrosis Transmembrane Conductance Regulator in the Endoplasmic Reticulum." The Journal of Biological Chemistry **286**(50): 43071-43080.

Bankevich, A., S. Nurk, *et al.* (2012). "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." Journal of Computational Biology **19**(5): 455-477.

Barkal, A. A., S. Srinivasan, *et al.* (2016). "Cas9 Functionally Opens Chromatin." PLoS ONE **11**(3): e0152683.

Barrangou, R., C. Fremaux, *et al.* (2007). "CRISPR provides acquired resistance against viruses in prokaryotes." Science **315**(5819): 1709-1712.

Baruteau, J., S. N. Waddington, *et al.* (2017). "Gene therapy for monogenic liver diseases: clinical successes, current challenges and future prospects." J Inherit Metab Dis.

Beck, S. E., L. A. Jones, *et al.* (1999). "Repeated delivery of adeno-associated virus vectors to the rabbit airway." J Virol **73**(11): 9446-9455.

Bengtsson, N. E., J. K. Hall, *et al.* (2017). "Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy." Nat Commun **8**: 14454.

Bhargava, R., D. O. Onyango, *et al.* (2016). "Regulation of Single Strand Annealing and its role in genome maintenance." Trends in genetics : TIG **32**(9): 566-575.

Boink, G. J. and R. B. Robinson (2014). "Gene therapy for restoring heart rhythm." J Cardiovasc Pharmacol Ther **19**(5): 426-438.

Bolotin, A., B. Quinquis, *et al.* (2005). "Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin." Microbiology **151**(Pt 8): 2551-2561.

Bolukbasi, M. F., A. Gupta, *et al.* (2015). "DNA-binding-domain fusions enhance the targeting range and precision of Cas9." Nat Meth **advance online publication**.

Booth, C., H. B. Gaspar, *et al.* (2016). "Treating Immunodeficiency through HSC Gene Therapy." Trends Mol Med **22**(4): 317-327.

Buckley, S. M., S. J. Howe, *et al.* (2008). "Lentiviral transduction of the murine lung provides efficient pseudotype and developmental stage-dependent cell-specific transgene expression." Gene Ther **15**(16): 1167-1175.

Bustamante-Marin, X. M. and L. E. Ostrowski (2017). "Cilia and Mucociliary Clearance." Cold Spring Harb Perspect Biol **9**(4).

Cao, J., L. Wu, *et al.* (2016). "An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting." Nucleic Acids Res.

Carpenter, A. E., T. R. Jones, *et al.* (2006). "CellProfiler: image analysis software for identifying and quantifying cell phenotypes." Genome Biology **7**(10): R100.

Casini, A., M. Olivieri, *et al.* (2018). "A highly specific SpCas9 variant is identified by in vivo screening in yeast." Nature Biotechnology.

Caspar, S. M., N. Dubacher, *et al.* (2017). "Clinical sequencing: from raw data to diagnosis with lifetime value." Clin Genet.

Castellani, S. and M. Conese (2010). "Lentiviral Vectors and Cystic Fibrosis Gene Therapy." Viruses **2**(2): 395-412.

Certo, M. T., B. Y. Ryu, *et al.* (2011). "Tracking genome engineering outcome at individual DNA breakpoints." Nat Methods. **8**.

Chen, J. S., Y. S. Dagdas, *et al.* (2017). "Enhanced proofreading governs CRISPR-Cas9 targeting accuracy." Nature.

Chen, S., B. Lee, *et al.* (2016). "Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes." J Biol Chem **291**(28): 14457-14467.

Chen, X., J. Liu, *et al.* (2017). "The Chromatin Structure Differentially Impacts High-Specificity CRISPR-Cas9 Nuclease Strategies." Mol Ther Nucleic Acids **8**: 558-563.

Chen, X., M. Rinsma, *et al.* (2016). "Probing the impact of chromatin conformation on genome editing tools." Nucleic Acids Res **44**(13): 6482-6492.

Chew, W. L., M. Tabebordbar, *et al.* (2016). "A multi-functional AAV-CRISPR-Cas9 and its host response." Nature methods **13**(10): 868-874.

Chhin, B., D. Negre, *et al.* (2009). "Ciliary Beating Recovery in Deficient Human Airway Epithelial Cells after Lentivirus Ex Vivo Gene Therapy." PLoS Genetics **5**(3): e1000422.

Chirmule, N., W. Xiao, *et al.* (2000). "Humoral immunity to adeno-associated virus type 2 vectors following administration to murine and nonhuman primate muscle." J Virol **74**(5): 2420-2425.

Choi, M., U. I. Scholl, *et al.* (2009). "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing." Proc Natl Acad Sci U S A **106**(45): 19096-19101.

Chu, V. T., T. Weber, *et al.* (2015). "Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells." Nat Biotechnol. **33**.

Clinical Lab Products. (2017, 27th April 2017). "What Is SMRT Sequencing?" Retrieved 6th February, 2018, from http://www.clpmag.com/2017/04/smrt-sequencing/.

Cong, L., F. A. Ran, *et al.* (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science. **339**.

Cox, D. B. T., R. J. Platt, *et al.* (2015). "Therapeutic Genome Editing: Prospects and Challenges." Nature medicine **21**(2): 121-131.

Crosetto, N., A. Mitra, *et al.* (2013). "Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing." Nat Methods **10**(4): 361-365.

Cutting, G. R. (2014). "Cystic fibrosis genetics: from molecular understanding to clinical application." Nature Reviews Genetics **16**: 45.

Damseh, N., N. Quercia, *et al.* (2017). "Primary ciliary dyskinesia: mechanisms and management." The Application of Clinical Genetics **10**: 67-74.

Daniels, M. L. A. and P. G. Noone (2015). "Genetics, diagnosis, and future treatment strategies for primary ciliary dyskinesia." Expert opinion on orphan drugs **3**(1): 31-44.

DeWitt, M., J. E. Corn, *et al.* (2017). "Genome editing via delivery of Cas9 ribonucleoprotein." Methods.

Doench, J. G., E. Hartenian, *et al.* (2014). "Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation." Nat Biotechnol.

Duan, J., G. Lu, *et al.* (2014). "Genome-wide identification of CRISPR/Cas9 off-targets in human genome." Cell Research **24**: 1009.

Dunbar, C. E., K. A. High, *et al.* (2018). "Gene therapy comes of age." Science **359**(6372).

Dupré, A., L. Boyer-Chatenet, *et al.* (2008). "A forward chemical genetic screen reveals an inhibitor of the Mre11–Rad50–Nbs1 complex." Nature chemical biology **4**(2): 119-125.

Eisenstein, M. (2017). "An ace in the hole for DNA sequencing." Nature **550**: 285.

El Refaey, M., L. Xu, *et al.* (2017). "In Vivo Genome Editing Restores Dystrophin Expression and Cardiac Function in Dystrophic Mice." Circ Res **121**(8): 923-929.

European Biotechnology. (2017, 21st April 2017). "Uniqure withdraws €1m drug Glybera from market." Retrieved 8th February, 2018, from https://european-biotechnology.com/up-to-date/latest-news/news/uniqure-withdraws-eur1m-drug-glybera-from-market.html.

European Organisation for Rare Diseases. (2005, November 2005). "Rare Diseases: understanding this Public Health Priority." Retrieved 5th February, 2018, from https://www.eurordis.org/IMG/pdf/princeps_document-EN.pdf.

Feng, Y., Y. Zhang, *et al.* (2015). "Nanopore-based Fourth-generation DNA Sequencing Technology." Genomics, Proteomics & Bioinformatics **13**(1): 4-16.

Franco, S., M. M. Murphy, *et al.* (2008). "DNA-PKcs and Artemis function in the end-joining phase of immunoglobulin heavy chain class switch recombination." The Journal of Experimental Medicine **205**(3): 557-564.

Friedland, A. E., R. Baral, *et al.* (2015). "Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications." Genome Biol **16**(1): 257.

Friedmann, T. and R. Roblin (1972). "Gene Therapy for Human Genetic Disease?" Science **175**(4025): 949.

Fu, Y., J. A. Foden, *et al.* (2013). "High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells." Nat Biotechnol **31**(9): 822-826.

Fu, Y., J. D. Sander, *et al.* (2014). "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs." Nat Biotechnol **32**(3): 279-284.

Gaj, T., B. E. Epstein, *et al.* (2015). "Genome engineering using adeno-associated virus: basic and clinical research applications." Mol Ther.

Gaj, T., C. A. Gersbach, *et al.* (2013). "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering." Trends Biotechnol **31**(7): 397-405.

Gaj, T., B. T. Staahl, *et al.* (2017). "Targeted gene knock-in by homology-directed genome editing using Cas9 ribonucleoprotein and AAV donor delivery." Nucleic Acids Res **45**(11): e98.

Gammie, T., C. Y. Lu, *et al.* (2015). "Access to Orphan Drugs: A Comprehensive Review of Legislations, Regulations and Policies in 35 Countries." PLoS One **10**(10): e0140002.

Garneau, J. E., M. E. Dupuis, *et al.* (2010). "The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA." Nature **468**(7320): 67-71.

Gaudelli, N. M., A. C. Komor, *et al.* (2017). "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage." Nature **551**: 464.

Gkotzamanidou, M., E. Terpos, *et al.* (2016). "DNA repair of myeloma plasma cells correlates with clinical outcome: the effect of the nonhomologous end-joining inhibitor SCR7." Blood **128**(9): 1214-1225.

Glaser, A., B. McColl, *et al.* (2016). "GFP to BFP Conversion: A Versatile Assay for the Quantification of CRISPR/Cas9-mediated Genome Editing." Mol Ther Nucleic Acids **5**: e334.

Goodwin, S., J. D. McPherson, *et al.* (2016). "Coming of age: ten years of next-generation sequencing technologies." Nat Rev Genet **17**(6): 333-351.

Gray-Edwards, H., A. N. Randle, *et al.* (2017). "AAV gene therapy in a sheep model of Tay-Sachs disease." Hum Gene Ther.

Griesenbach, U., M. Inoue, *et al.* (2012). "Assessment of F/HN-pseudotyped lentivirus as a clinically relevant vector for lung gene therapy." Am J Respir Crit Care Med **186**(9): 846-856.

Guggino, W. B. and L. Cebotaru (2017). "Adeno-Associated Virus (AAV) gene therapy for cystic fibrosis: current barriers and recent developments." Expert Opin Biol Ther **17**(10): 1265-1273.

Hadjantonakis, A. K. and V. E. Papaioannou (2004). "Dynamic in vivo imaging and cell tracking using a histone fluorescent protein fusion in mice." BMC Biotechnol **4**: 33.

Haeussler, M., K. Schönig, *et al.* (2016). "Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR." Genome Biology **17**: 148.

Hendel, A., E. J. Fine, *et al.* (2015). "Quantifying On and Off-Target Genome Editing." Trends in biotechnology **33**(2): 132-140.

Hendel, A., E. J. Kildebeck, *et al.* (2014). "Quantifying Genome Editing Outcomes at Endogenous Loci using SMRT Sequencing." Cell reports **7**(1): 293-305.

Horani, A., T. W. Ferkol, *et al.* (2016). "Genetics and biology of primary ciliary dyskinesia." Paediatr Respir Rev **18**: 18-24.

Horlbeck, M. A., L. B. Witkowsky, *et al.* (2016). "Nucleosomes impede Cas9 access to DNA and." Elife **5**.

Hruscha, A., P. Krawitz, *et al.* (2013). "Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish." Development (Cambridge) **140**(24): 4982-4987.

Hsu, P. D., D. A. Scott, *et al.* (2013). "DNA targeting specificity of RNA-guided Cas9 nucleases." Nat Biotech **31**(9): 827-832.

Hu, Z., Z. Shi, *et al.* (2018). "Ligase IV inhibitor SCR7 enhances gene editing directed by CRISPR-Cas9 and ssODN in human cancer cells." Cell Biosci **8**: 12.

Huang, L., F. Ma, *et al.* (2015). "Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications." Annu Rev Genomics Hum Genet **16**: 79-102.

Hutt, D. M., D. M. Roth, *et al.* (2012). "FK506 binding protein 8 peptidylprolyl isomerase activity manages a late stage of cystic fibrosis transmembrane conductance regulator (CFTR) folding and stability." J Biol Chem **287**(26): 21914-21925.

Hwang, W. Y., Y. Fu, *et al.* (2013). "Heritable and precise zebrafish genome editing using a CRISPR-Cas system." PLoS One **8**(7): e68708.

Hwang, W. Y., Y. Fu, *et al.* (2013). "Efficient genome editing in zebrafish using a CRISPR-Cas system." Nature Biotechnology **31**: 227.

Isaac, R. S., F. Jiang, *et al.* (2016). "Nucleosome breathing and remodeling constrain CRISPR-Cas9 function." Elife **5**.

Ishizu, T., S. Higo, *et al.* (2017). "Targeted Genome Replacement via Homology-directed Repair in Non-dividing Cardiomyocytes." Sci Rep **7**(1): 9363.

Jain, M., H. E. Olsen, *et al.* (2016). "The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community." Genome Biology **17**(1): 239.

Jansen, R., J. D. Embden, *et al.* (2002). "Identification of genes that are associated with DNA repeats in prokaryotes." Mol Microbiol **43**(6): 1565-1575.

Jensen, K. T., L. Floe, *et al.* (2017). "Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency." FEBS Lett.

Jinek, M., K. Chylinski, *et al.* (2012). "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity." Science **337**(6096): 816-821.

Jones, T. R., I. H. Kang, *et al.* (2008). "CellProfiler Analyst: data exploration and analysis software for complex image-based screens." BMC Bioinformatics **9**(1): 482.

Kakarougkas, A. and P. A. Jeggo (2014). "DNA DSB repair pathway choice: an orchestrated handover mechanism." Br J Radiol **87**(1035): 20130685.

Kanda, T., K. F. Sullivan, *et al.* (1998). "Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells." Curr Biol **8**(7): 377-385.

Kay, M. A., C. Y. He, *et al.* (2010). "A robust system for production of minicircle DNA vectors." Nat Biotechnol **28**(12): 1287-1289.

Kim, D., S. Bae, *et al.* (2015). "Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells." Nature Methods **12**: 237.

Kim, Y. B., A. C. Komor, *et al.* (2017). "Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions." Nat Biotechnol **35**(4): 371-376.

Kleinstiver, B. P., V. Pattanayak, *et al.* (2016). "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." Nature.

Koboldt, D. C., Q. Zhang, *et al.* (2012). "VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing." Genome Res **22**(3): 568-576.

Kolodziejczyk, A. A., J. K. Kim, *et al.* (2015). "The technology and biology of single-cell RNA sequencing." Mol Cell **58**(4): 610-620.

Komor, A. C., Y. B. Kim, *et al.* (2016). "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage." Nature **533**(7603): 420-424.

Koren, S., B. P. Walenz, *et al.* (2017). "Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation." Genome Research **27**(5): 722-736.

Koressaar, T. and M. Remm (2007). "Enhancements and modifications of primer design program Primer3." Bioinformatics **23**(10): 1289-1291.

Kosicki, M., S. S. Rajan, *et al.* (2017). "Dynamics of Indel Profiles Induced by Various CRISPR/Cas9 Delivery Methods." Prog Mol Biol Transl Sci **152**: 49-67.

Kowalczykowski, S. C. (2015). "An Overview of the Molecular Mechanisms of Recombinational DNA Repair." Cold Spring Harbor Perspectives in Biology **7**(11): a016410.

Kuhar, R., K. S. Gwiazda, *et al.* (2014). "Novel fluorescent genome editing reporters for monitoring DNA repair pathway utilization at endonuclease-induced breaks." Nucleic Acids Research **42**(1): e4-e4.

Kurkowiak, M., E. Zietkiewicz, *et al.* (2016). "ZMYND10 - Mutation Analysis in Slavic Patients with Primary Ciliary Dyskinesia." PLoS One **11**(1): e0148067.

Kurosaki, F., R. Uchibori, *et al.* (2017). "Optimization of adeno-associated virus vector-mediated gene transfer to the respiratory tract." Gene Ther.

Kuscu, C., S. Arslan, *et al.* (2014). "Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease." Nat Biotechnol.

Lai, M., M. Pifferi, *et al.* (2016). "Gene editing of DNAH11 restores normal cilia motility in primary ciliary dyskinesia." Journal of Medical Genetics.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**.

Lau, C. H. and Y. Suh (2017). "In vivo genome editing in animals using AAV-CRISPR system: applications to translational research of human disease." F1000Res **6**: 2153.

Leahy, J. J., B. T. Golding, *et al.* (2004). "Identification of a highly potent and selective DNA-dependent protein kinase (DNA-PK) inhibitor (NU7441) by screening of chromenone libraries." Bioorg Med Chem Lett. **14**.

Lessard, S., L. Francioli, *et al.* (2017). "Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci." Proc Natl Acad Sci U S A **114**(52): E11257-e11266.

Li, D., Z. Qiu, *et al.* (2013). "Heritable gene targeting in the mouse and rat using a CRISPR-Cas system." Nat Biotechnol **31**(8): 681-683.

Li, G., X. Zhang, *et al.* (2017). "Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells." Sci Rep **7**(1): 8943.

Li, H., B. Handsaker, *et al.* (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Li, J., W. Sun, *et al.* (2008). "Protein trans-splicing as a means for viral vector-mediated in vivo gene therapy." Hum Gene Ther **19**(9): 958-964.

Li, K., G. Wang, *et al.* (2014). "Optimization of Genome Engineering Approaches with the CRISPR/Cas9 System." PLoS ONE **9**(8): e105779.

Li, X. and W.-D. Heyer (2008). "Homologous recombination in DNA repair and DNA damage tolerance." Cell Research **18**: 99.

Lieber, M. R. (2008). "The mechanism of human nonhomologous DNA end joining." J Biol Chem **283**(1): 1-5.

Lieber, M. R. (2010). "The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway." Annu Rev Biochem. **79**.

Lin, S., B. T. Staahl, *et al.* (2014). "Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery." eLife. **3**.

Lin, S. R., H. C. Yang, *et al.* (2014). "The CRISPR/Cas9 System Facilitates Clearance of the Intrahepatic HBV Templates In Vivo." Mol Ther Nucleic Acids **3**: e186.

Lisowski, L., A. P. Dane, *et al.* (2014). "Selection and evaluation of clinically relevant AAV variants in a xenograft liver model." Nature **506**(7488): 382-386.

Liu, K. I., M. N. Ramli, *et al.* (2016). "A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing." Nat Chem Biol.

Loman, N. J. and A. R. Quinlan (2014). "Poretools: a toolkit for analyzing nanopore sequence data." Bioinformatics **30**(23): 3399-3401.

Lu, H., F. Giordano, *et al.* (2016). "Oxford Nanopore MinION Sequencing and Genome Assembly." Genomics Proteomics Bioinformatics **14**(5): 265-279.

Lu, J., C. Zhao, *et al.* (2018). "Multimode drug inducible CRISPR/Cas9 devices for transcriptional activation and genome editing." Nucleic Acids Research **46**(5): e25-e25.

Luo, D. and W. M. Saltzman (2000). "Synthetic DNA delivery systems." Nat Biotechnol **18**(1): 33-37.

Ma, H., N. Marti-Gutierrez, *et al.* (2017). "Correction of a pathogenic gene mutation in human embryos." Nature.

Mali, G. R., P. Yeyati, *et al.* (2017). "ZMYND10 functions in a chaperone relay during axonemal dynein assembly." bioRxiv.

Mali, P., L. Yang, *et al.* (2013). "RNA-Guided Human Genome Engineering via Cas9." Science **339**(6121): 823-826.

Marraffini, L. A. and E. J. Sontheimer (2008). "CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA." Science **322**(5909): 1843-1845.

Marsit, C. J., D. H. Kim, *et al.* (2005). "Hypermethylation of RASSF1A and BLU tumor suppressor genes in non-small cell lung cancer: implications for tobacco smoking during adolescence." Int J Cancer **114**(2): 219-223.

Martino, A. T., E. Basner-Tschakarjan, *et al.* (2013). "Engineered AAV vector minimizes in vivo targeting of transduced hepatocytes by capsid-specific CD8+ T cells." Blood **121**(12): 2224-2233.

Maruyama, T., S. K. Dougan, *et al.* (2015). "Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining." Nat Biotechnol. **33**.

McIntyre, J. C., E. E. Davis, *et al.* (2012). "Gene therapy rescues cilia defects and restores olfactory function in a mammalian ciliopathy model." Nat Med **18**(9): 1423-1428.

McKenna, A., M. Hanna, *et al.* (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

McVey, M. and S. E. Lee (2008). "MMEJ repair of double-strand breaks (director’s cut): deleted sequences and alternative endings." Trends in Genetics **24**(11): 529-538.

Meienberg, J., R. Bruggmann, *et al.* (2016). "Clinical sequencing: is WGS the better WES?" Hum Genet **135**(3): 359-362.

Mir, A., A. Edraki, *et al.* (2018). "Type II-C CRISPR-Cas9 Biology, Mechanism, and Application." ACS Chemical Biology **13**(2): 357-365.

Mirra, V., C. Werner, *et al.* (2017). "Primary Ciliary Dyskinesia: An Update on Clinical Aspects, Genetics, Diagnosis, and Future Treatment Strategies." Front Pediatr **5**: 135.

Mitchison, H. M. and E. M. Valente (2017). "Motile and non-motile cilia in human pathology: from function to phenotypes." The Journal of Pathology **241**(2): 294-309.

Mitomo, K., U. Griesenbach, *et al.* (2010). "Toward gene therapy for cystic fibrosis using a lentivirus pseudotyped with Sendai virus envelopes." Mol Ther **18**(6): 1173-1182.

Miyaoka, Y., J. R. Berman, *et al.* (2016). "Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing." Sci Rep **6**: 23549.

Miyoshi, H., U. Blömer, *et al.* (1998). "Development of a Self-Inactivating Lentivirus Vector." Journal of Virology **72**(10): 8150-8157.

Mojica, F. J., C. Diez-Villasenor, *et al.* (2009). "Short motif sequences determine the targets of the prokaryotic CRISPR defence system." Microbiology **155**(Pt 3): 733-740.

Mojica, F. J., C. Diez-Villasenor, *et al.* (2005). "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements." J Mol Evol **60**(2): 174-182.

Moore, Daniel J., A. Onoufriadis, *et al.* (2013). "Mutations in ZMYND10, a Gene Essential for Proper Axonemal Assembly of Inner and Outer Dynein Arms in Humans and Flies, Cause Primary Ciliary Dyskinesia." American Journal of Human Genetics **93**(2): 346-356.

Mottais, A., T. Le Gall, *et al.* (2017). "Enhancement of lung gene delivery after aerosol: a new strategy using non-viral complexes with antibacterial properties." Biosci Rep **37**(6).

Munye, M. M., A. D. Tagalakis, *et al.* (2016). "Minicircle DNA Provides Enhanced and Prolonged Transgene Expression Following Airway Gene Transfer." Scientific Reports **6**: 23125.

Muzumdar, M. D., B. Tasic, *et al.* (2007). "A global double-fluorescent Cre reporter mouse." Genesis **45**(9): 593-605.

Nagarajan, N. and M. Pop (2013). "Sequence assembly demystified." Nat Rev Genet **14**(3): 157-167.

Nelson, C. E., C. H. Hakim, *et al.* (2016). "In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy." Science.

Next Gen Seek. (2016, 8th March 2016). "Oxford Nanopore Updates: Reveals the nanopore used in Nanopore devices." Retrieved 6th February, 2018, from http://nextgenseek.com/2016/03/oxford-nanopore-updates-from-technology-focus-by-clive-g-brown/.

Novak, A., C. Guo, *et al.* (2000). "Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon Cre-mediated excision." Genesis **28**(3-4): 147-155.

O'Geen, H., I. M. Henry, *et al.* (2015). "A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture." Nucleic Acids Res **43**(6): 3389-3404.

Ostrowski, L. E., W. Yin, *et al.* (2014). "Restoring ciliary function to differentiated primary ciliary dyskinesia cells with a lentiviral vector." Gene Ther **21**(3): 253-261.

Paquet, D., D. Kwart, *et al.* (2016). "Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9." Nature **533**(7601): 125-129.

Perez, A. R., Y. Pritykin, *et al.* (2017). "GuideScan software for improved single and paired CRISPR guide RNA design." Nature Biotechnology **35**: 347.

Pinder, J., J. Salsman, *et al.* (2015). "Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing." Nucleic Acids Res **43**(19): 9379-9392.

Pinello, L., M. C. Canver, *et al.* (2016). "Analyzing CRISPR genome-editing experiments with CRISPResso." Nat Biotech **34**(7): 695-697.

Pogue, R. E., D. P. Cavalcanti, *et al.* (2018). "Rare genetic diseases: update on diagnosis, treatment and online resources." Drug Discov Today **23**(1): 187-195.

Polineni, D., S. D. Davis, *et al.* (2015). "Treatment recommendations in Primary Ciliary Dyskinesia." Paediatr Respir Rev **18**: 39-45.

Pourcel, C., G. Salvignol, *et al.* (2005). "CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies." Microbiology **151**(Pt 3): 653-663.

Qasim, W., H. Zhan, *et al.* (2017). "Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells." Sci Transl Med **9**(374).

Quail, M. A., M. Smith, *et al.* (2012). "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers." BMC Genomics **13**.

Ramanan, V., A. Shlomai, *et al.* (2015). "CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus." Sci Rep **5**: 10833.

Ran, F. A., P. D. Hsu, *et al.* (2013). "Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity." Cell **154**(6): 1380-1389.

Ran, F. A., P. D. Hsu, *et al.* (2013). "Genome engineering using the CRISPR-Cas9 system." Nat Protoc **8**(11): 2281-2308.

Reiter, J. F. and M. R. Leroux (2017). "Genes and molecular pathways underpinning ciliopathies." Nat Rev Mol Cell Biol **18**(9): 533-547.

Richardson, C. D., G. J. Ray, *et al.* (2016). "Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA." Nat Biotechnol.

Riken Center for Developmental Biology. (2017, 11th January 2017). "HITI: An innovative in vivo genome-editing technology." Retrieved 19th June, 2018, from http://www.cdb.riken.jp/en/news/2017/researches/0111_10017.html.

Rio, P., R. Banos, *et al.* (2014). "Targeted gene therapy and cell reprogramming in Fanconi anemia." EMBO Mol Med **6**(6): 835-848.

Robert, F., M. Barbeau, *et al.* (2015). "Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing." Genome Med **7**: 93.

Robinson, J. T., H. Thorvaldsdóttir, *et al.* (2011). "Integrative genomics viewer." Nature Biotechnology **29**: 24.

Rueden, C. T., J. Schindelin, *et al.* (2017). "ImageJ2: ImageJ for the next generation of scientific image data." BMC Bioinformatics **18**(1): 529.

Sander, J. D., M. L. Maeder, *et al.* (2010). "ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool." Nucleic Acids Res **38**(Web Server issue): W462-468.

Sander, J. D., P. Zaback, *et al.* (2007). "Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool." Nucleic Acids Research **35**(Web Server issue): W599-W605.

Schindelin, J., I. Arganda-Carreras, *et al.* (2012). "Fiji: an open-source platform for biological-image analysis." Nature Methods **9**: 676.

Senior, M. (2017). "After Glybera's withdrawal, what's next for gene therapy?" Nature Biotechnology **35**: 491.

Senís, E., C. Fatouros, *et al.* (2014). "CRISPR/Cas9-mediated genome engineering: An adeno-associated viral (AAV) vector toolbox." Biotechnology Journal: n/a-n/a.

Senturk, S., N. H. Shirole, *et al.* (2017). "Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization." Nature Communications **8**: 14370.

Seol, J. H., E. Y. Shim, *et al.* (2017). "Microhomology-mediated end joining: Good, bad and ugly." Mutat Res.

Sfeir, A. and L. S. Symington (2015). "Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?" Trends Biochem Sci **40**(11): 701-714.

Shao, S., C. Ren*, et al.* (2017). "Enhancing CRISPR/Cas9-mediated homology-directed repair in mammalian cells by expressing Saccharomyces cerevisiae Rad52." Int J Biochem Cell Biol **92**: 43-52.

Sharma, S., S. M. Javadekar*, et al.* (2015). "Homology and enzymatic requirements of microhomology-dependent alternative end joining." Cell Death & Disease **6**(3): e1697.

Slaymaker, I. M., L. Gao*, et al.* (2015). "Rationally engineered Cas9 nucleases with improved specificity." Science.

Smith, C., A. Gore*, et al.* (2014). "Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs." Cell Stem Cell **15**(1): 12-13.

Soden, S. E., C. J. Saunders*, et al.* (2014). "Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders." Sci Transl Med **6**(265): 265ra168.

Song, J., D. Yang*, et al.* (2016). "RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency." Nat Commun **7**.

Sovic, I., M. Sikic*, et al.* (2016). "Fast and sensitive mapping of nanopore sequencing reads with GraphMap." Nat Commun **7**: 11307.

Spassky, N. and A. Meunier (2017). "The development and functions of multiciliated epithelia." Nat Rev Mol Cell Biol.

Srinivas, S., T. Watanabe*, et al.* (2001). "Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus." BMC Developmental Biology **1**(1): 4.

Srivastava, M., M. Nambiar*, et al.* (2012). "An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression." Cell **151**(7): 1474-1487.

Stavropoulos, D. J., D. Merico*, et al.* (2016). "Whole Genome Sequencing Expands Diagnostic Utility and Improves Clinical Management in Pediatric Medicine." NPJ Genom Med **1**.

Steines, B., D. D. Dickey, *et al.* (2016). "CFTR gene transfer with AAV improves early cystic fibrosis pig phenotypes." JCI Insight **1**(14): e88728.

Stella, S. and G. Montoya (2016). "The genome editing revolution: A CRISPR-Cas TALE off-target story." Bioessays **38 Suppl 1**: S4-s13.

Suda, T. and D. Liu (2007). "Hydrodynamic gene delivery: its principles and applications." Mol Ther **15**(12): 2063-2069.

Sung, Y. H., J. M. Kim, *et al.* (2014). "Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases." Genome Res **24**(1): 125-131.

Suzuki, K. and J. C. Izpisua Belmonte (2018). "In vivo genome editing via the HITI method as a tool for gene therapy." Journal of Human Genetics **63**(2): 157-164.

Suzuki, K., Y. Tsunekawa, *et al.* (2016). "In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration." Nature **540**: 144.

Tang, L., Y. Zeng, *et al.* (2017). "CRISPR/Cas9-mediated gene editing in human zygotes using Cas9 protein." Mol Genet Genomics.

Tebas, P., D. Stein, *et al.* (2014). "Gene Editing of CCR5 in Autologous CD4 T Cells of Persons Infected with HIV." New England Journal of Medicine **370**(10): 901-910.

Thermo Fisher Scientific. (2015). "User Guide: GeneArt™ Platinum™ Cas9 Nuclease." Retrieved 9th January, 2017, from https://assets.thermofisher.com/TFS-Assets/LSG/manuals/geneart_platinum_cas9_nuclease_man.pdf.

Thermo Fisher Scientific. (2016). "User Guide: GeneArt Precision gRNA Synthesis Kit." Retrieved 9th January, 2017, from https://assets.thermofisher.com/TFS-Assets/LSG/manuals/geneart_precision_gRNA_synthesis_kit_man.pdf.

Thorvaldsdottir, H., J. T. Robinson, *et al.* (2013). "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration." Brief Bioinform **14**(2): 178-192.

Touchot, N. and M. Flume (2017). "Early Insights from Commercialization of Gene Therapies in Europe." Genes **8**(2): 78.

Tsai, S. Q., N. T. Nguyen, *et al.* (2017). "CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets." Nat Methods **14**(6): 607-614.

Tsai, S. Q., Z. Zheng, *et al.* (2014). "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases." Nature Biotechnology **33**: 187.

Tsouroula, K., A. Furst, *et al.* (2016). "Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin." Mol Cell **63**(2): 293-305.

Tu, Z., W. Yang, *et al.* (2017). "Promoting Cas9 degradation reduces mosaic mutations in non-human primate embryos." Scientific Reports **7**: 42081.

U.S. Food and Drug Administration. (2017, 19th December 2017). "Developing Products for Rare Diseases & Conditions." Retrieved 5th February, 2018, from https://www.fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/default.htm.

UK Department of Health. (2013). "The UK Strategy for Rare Diseases." Retrieved 5th February, 2018, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/260562/UK_Strategy_for_Rare_Diseases.pdf.

Untergasser, A., I. Cutcutache, *et al.* (2012). "Primer3—new capabilities and interfaces." Nucleic Acids Research **40**(15): e115-e115.

Veres, A., Bridget S. Gosis, *et al.* (2014). "Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing." Cell Stem Cell **15**(1): 27-30.

Vidović, D., M. S. Carlon, *et al.* (2015). "rAAV-CFTRΔR Rescues the Cystic Fibrosis Phenotype in Human Intestinal Organoids and Cystic Fibrosis Mice." American Journal of Respiratory and Critical Care Medicine **193**(3): 288-298.

Wang, D., H. Fischer, *et al.* (1999). "Efficient CFTR expression from AAV vectors packaged with promoters--the second generation." Gene Ther **6**(4): 667-675.

Wang, H., H. Yang, *et al.* (2013). "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering." Cell. **153**.

Wang, H., H. Yang, *et al.* (2013). "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering." Cell **153**(4): 910-918.

Watts, F. Z. (2016). "Repair of DNA Double-Strand Breaks in Heterochromatin." Biomolecules **6**(4).

Wells, D. J. (2017). "Systemic AAV Gene Therapy Close to Clinical Trials for Several Neuromuscular Diseases." Mol Ther.

Wen, Y., G. Liao, *et al.* (2017). "A Stable but Reversible Integrated Surrogate Reporter for Assaying CRISPR/Cas9-Stimulated Homology-directed Repair." J Biol Chem.

Weymann, D., J. Laskin, *et al.* (2017). "The cost and cost trajectory of whole-genome analysis guiding treatment of patients with advanced cancers." Molecular Genetics & Genomic Medicine **5**(3): 251-260.

Wick, R. R., M. B. Schultz, *et al.* (2015). "Bandage: interactive visualization of de novo genome assemblies." Bioinformatics **31**(20): 3350-3352.

Wu, X., D. A. Scott, *et al.* (2014). "Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells." Nature Biotechnology **32**: 670.

Wu, Y., D. Liang, *et al.* (2013). "Correction of a genetic disease in mouse via use of CRISPR-Cas9." Cell Stem Cell **13**(6): 659-662.

Xie, C., Y. P. Zhang, *et al.* (2016). "Genome editing with CRISPR/Cas9 in postnatal mice corrects PRKAG2 cardiac syndrome." Cell Res **26**(10): 1099-1111.

Yan, Z., Z. Feng, *et al.* (2017). "Human Bocavirus Type-1 Capsid Facilitates the Transduction of Ferret Airways by Adeno-Associated Virus Genomes." Hum Gene Ther.

Yang, D., M. A. Scavuzzo, *et al.* (2016). "Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases." Sci Rep **6**: 21264.

Yang, H., H. Wang, *et al.* (2013). "One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering." Cell. **154**.

Yau, W. L., H. L. Lung, *et al.* (2006). "Functional studies of the chromosome 3p21.3 candidate tumor suppressor gene BLU/ZMYND10 in nasopharyngeal carcinoma." Int J Cancer **119**(12): 2821-2826.

Yin, H., W. Xue, *et al.* (2014). "Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype." Nat Biotechnol **32**(6): 551-553.

Yu, C., Y. Liu, *et al.* (2015). "Small molecules enhance CRISPR genome editing in pluripotent stem cells." Cell Stem Cell. **16**.

Yu, W., S. Mookherjee, *et al.* (2017). "Nrl knockdown by AAV-delivered CRISPR/Cas9 prevents retinal degeneration in mice." Nat Commun **8**: 14716.

Zaboikin, M., T. Zaboikina, *et al.* (2017). "Non-Homologous End Joining and Homology Directed DNA Repair Frequency of Double-Stranded Breaks Introduced by Genome Editing Reagents." PLoS One **12**(1): e0169931.

Zariwala, Maimoona A., Heon Y. Gee, *et al.* (2013). "ZMYND10 Is Mutated in Primary Ciliary Dyskinesia and Interacts with LRRC6." American Journal of Human Genetics **93**(2): 336-345.

Zariwala, M. A., H. Y. Gee, *et al.* (2013). "ZMYND10 is mutated in primary ciliary dyskinesia and interacts with LRRC6." Am J Hum Genet **93**(2): 336-345.

Zhang, J. P., X. L. Li, *et al.* (2017). "Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage." Genome Biol **18**(1): 35.

Zhang, J. P., X. L. Li, *et al.* (2016). "Different Effects of sgRNA Length on CRISPR-mediated Gene Knockout Efficiency." Sci Rep **6**: 28566.

Zhang, L., D. Wang, *et al.* (1998). "Efficient expression of CFTR function with adeno-associated virus vectors that carry shortened CFTR genes." Proceedings of the National Academy of Sciences of the United States of America **95**(17): 10158-10163.

Zhang Lab, M. (2017). "CRISPR Design." Retrieved 18th February, 2015, from http://crispr.mit.edu/.

Zhao, M. and D. Q. Wei (2017). "Rare Diseases: Drug Discovery and Informatics Resource." Interdiscip Sci.

Zhen, S., L. Hua, *et al.* (2015). "Harnessing the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated Cas9 system to disrupt the hepatitis B virus." Gene Ther **22**(5): 404-412.

Zinc Finger Consortium. (2017). "ZiFiT Targeter." Retrieved 18th February, 2015, from http://zifit.partners.org/ZiFiT/.

Zufferey, R., T. Dull, *et al.* (1998). "Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery." Journal of Virology **72**(12): 9873-9880.